ABSTRACT

Title of dissertation: REPRESENTATIONAL CONTENT AND THE SCIENCE OF VISION

John Brendan Welsh Ritchie
Doctor of Philosophy, 2015

Dissertation directed by: Professor Peter Carruthers
Department of Philosophy

The general topic of my thesis is how vision science explains what we see, and how we see it. There are two themes often found in the explanations of vision science that I focus on. The first is the *Distal Object Thesis*: the internal representations that underlie object vision represent properties of entities in the distal world. The second is the *Transformational Thesis*: the function of the vision system is to transform information that is latent in the retinal image into a representational format that makes it available for use by further perceptual or cognitive systems. The ultimate aim of my project is to show that these two themes are in tension, and to suggest how the tension may be resolved.

The tension between these themes is, I argue, a result of their conflicting implications regarding the role of representational content (what a representation is "about") in the explanations of vision science. On the one hand, the Distal Object Thesis entails that the internal representations that underlie object vision qualify as a form of *mental* representation, and reflect a sense in which visual perception is indeed “objective”. Furthermore, I argue at length that a commitment to the Distal Object Thesis (and its consequences) is well-founded: mental representations are indeed an indispensable posit for explanations of aspects of object vision. On the other hand, the Transformational Thesis rests on the presupposition that the content of the internal representations in the visual system are fixed by a causally reliable, information carrying relation. The tension arises because carrying information is insufficient for fixing the content of mental representations. Thus the explanations of object vision that assume the Transformational Thesis, but require a commitment the Distal Object Thesis, are seemingly inadequate. Fortunately, some philosophical theories of *intentional content*, or the “aboutness” of mental representations, offer some strategies for reconciling these two themes in the explanations of vision science.

REPRESENTATIONAL CONTENT
AND THE SCIENCE OF VISION

by

John Brendan Welsh Ritchie

Dissertation submitted to the Faculty of the Graduate School of University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2015

Advisory Committee:
Professor Peter Carruthers, Chair
Professor Georges Rey
Professor Lindley Darden
Professor Gualtiero Piccinini, University of Missouri, St.Louis
Professor Michael Dougherty, Psychology

Dedication

To Sean—

También soy escritura

Y en este mismo instante

Alguien me deletrea

Acknowledgments

As I was completing this thesis, I found myself thinking again (and again) of the well-known passage from Eliot's *Little Gidding*:

We shall not cease from exploration
And the end of all our exploring
Will be to arrive where we started
And know the place for the first time.

Somehow this seems descriptive of my dissertation writing experience: as exploratory, and cyclical. And whatever modest forward progress I ultimately did achieve, it was due to the exertion of many external forces, who are all due thanks. Writing this thesis coincided with some of the most difficult periods of my life. I believe that, quite literally, I could not have done it without their help.

My first and greatest debt of thanks is to my thesis advisor, Peter Carruthers. It is an understatement to say that the path I took with my studies was circuitous, and it was almost entirely through areas that were not his primary expertise. Yet despite this, he showed unbelievable patience as I stumbled forward, while providing consistent (and persistent) encouragement. Working with him made me both a better philosopher and cognitive scientist. I could not have asked for a better advisor, and I consider it a great privilege to have been his student.

Next, I must thank Thomas Carlson, who recruited me into his lab shortly after he came to the University of Maryland. Tom gave me the opportunity to try my hand at experimental research, and introduced me to the splendor of vision science. Even though I was never his graduate student in any official capacity he nonetheless provided incredible mentorship and moral support throughout my studies. Similarly, although he was not officially involved with my dissertation, it was substantially shaped by his influence. I am grateful to him for the great trust he has shown in my scientific abilities.

Thanks are also due to the members of my dissertation committee. First off, Georges Rey was a constant source of inspiration during my studies at Maryland. In virtually all parts of the thesis I can see the influence of his ideas. I owe a debt of thanks to Lindley Darden, who taught me the importance of the history *and* philosophy of science, and to Michael Dougherty (my dean's representative) who taught me a great deal about the (philosophical) importance of statistics. Thanks are also in order to Gualtiero Piccinini who took the time to Skype in for the defense, and whose ideas also substantially influenced the project. And although he was not ultimately on the committee, I am deeply grateful to Aidan Lyon, who worked very hard to help me improve my writing.

My tenure at the University of Maryland was a long one, and I am grateful to the Department of Philosophy for tolerating my presence for so long. Louise Gilman deserves special thanks for her doting attention, while I was part of her flock of graduate students. I am also thankful to my cohort members: Lisa Leininger, Steve Frank Emet, Max Heiber, and Mark Engelbert, for their continued friendship throughout my time in graduate school. While pursuing my studies I was also lucky enough to live with a good number of housemates who provided companionship and moral support. Chief among them are Lucas Dunlap and Brock Rough, who provided much needed encouragement when my self-confidence was in short supply. When it came to the actual contents of the thesis, thanks are due to Dimiter Kirilov, Max Bialek, Evan Westra, and Andrew Knoll for many helpful conversations and suggestions. I also benefited from the friendship of individuals outside of the philosophy graduate program at Maryland—in particular, Alexis Wellwood in linguistics, and Bryce Huebner from Georgetown University. Lastly, I am indebted to Chris Vogel, for more reasons that I can enumerate here. Since we started the program together, he has been a constant companion. I could not have asked for a better friend with whom to share the ups and downs of graduate school.

This thesis was almost entirely (re)written while I was a research associate in the Department of Cognitive Science, at Macquarie University, in Sydney, Australia. I am thankful to Mark Williams and Anina Rich who gave me the opportunity to train as a cognitive neuroscientist, while I tried to write my long overdue dissertation. For being great colleagues and friends, thanks to Lincoln Colling, Glenn Carruthers, Alex Woolgar, Kiley Seymour, and Julia Misersky. Thanks also to Max Coltheart, David Kaplan, and especially, Colin Klein, who all provided wonderful feedback while the thesis was taking shape. Most of all, thanks to Susan Wardle, who was my partner in crime while I was at Macquarie, and who taught me the subtle art of overcoming dissertation inertia.

Thanks is due to my family and friends. I owe a special thanks to my mother, Mavis, and siblings, Aidan and Marlise, as well as my father, Craig. Their love and support made all the difference while I was pursuing my degree. Thanks are also due to Jing Chen, for her encouragement when I started my thesis. For the better part of the past year finishing this project was something of an all consuming obsession, and throughout that time Becka Clare kept me grounded. I am so grateful for her support, which has meant more to me than she can ever know.

Finally, I must thank my twin brother, Sean, who this dissertation is dedicated to. He is my better, as a person, poet, and philosopher. My debt of gratitude to him is enormous, and my only hope of repaying it is to live up to his example.

Table of Contents

List of Figures

Chapter 1: Questions of Content in Vision

1.1 Telling the Story of Vision

The general topic of this thesis is how vision science explains what we see, and how we see it. There is after all a main narrative to what we may call the *story* of vision. From the light that hits the retina information about the visible world is extracted, and then transformed, as it is carried through multiple stages of processing along the visual pathways of the brain. We see because each of these stages constitutes a collection of internal representations that make explicit information about different signals pulled from the patterns of electromagnetic radiation that reach the eyes. Representations of edge and orientation, tuned to the 2D structure of the retinal image, give way to those for shape, color, and motion, so that by the end we perceive something coherent, concrete and whole: a world of objects and events, bathed in light.

Or so the story goes. This narrative is the core of what I will call the *information-processing framework* in vision science, which is the dominant research program in the field (Marr, 1982; Palmer, 1999). There are two common themes to this narrative that I wish to focus on:

> *The Distal Object Thesis*: the end-stage internal representations that underlie object vision represent properties of entities in the distal world.

> *The Transformational Thesis*: the function of the visual system is to reformat ("transform") information latent in the retinal image into a representational format that makes it available for use by further perceptual or cognitive systems.

I am interested in these themes because of what they imply about the role of representational content in the story of vision. The ultimate aim of my project is to show that these two common themes to the information-processing narrative are in tension, and to spell out how the tension might be resolved.

On the one hand, the Distal Object Thesis entails that the internal representations that underlie object vision qualify as a form of *mental* representation, and reflect a sense in which visual perception is indeed "objective" (Burge, 2010). Furthermore, I will argue that these consequences of the thesis are well-founded; that is, mental representations—or rather *perceptual* representations—are an indispensable posit in explanations of object vision. On the other, I show that the Transformational Thesis rests on the presupposition that the content of the internal representations in the visual system are fixed by a causally reliable, information carrying relation. The tension arises because carrying information is insufficient for fixing the content of mental representations (Fodor, 1984). Thus explanations that rest on the Transformational Thesis will be inadequate to explain facts that depend on the Distal Object Thesis. Fortunately, certain philosophical theories of *inten-*

tional content, or the “aboutness” of mental representations, offer some strategies for reconciling these two themes in the story of vision.

1.2 Background and Motivation

Before I summarize the discussion to follow, let me provide some philosophical context for my project. Specifically, I will discuss the notions of mental representation, intentional content, and how philosophical discussion of these notions are related to the explanatory practices of cognitive science more generally.

One of the most salient features of the mind is that it exhibits *intentionality*: mental states are directed towards, or “about”, entities, properties or states of affairs. While ‘intentionality’ is a philosopher’s term of art, the idea that it picks out is quite familiar. For example, the various mental states we attribute to ourselves and others all appear to exhibit intentionality, including attitudes such as beliefs and desires, and perceptual states such as seeing or hearing. When I believe the day is sunny, desire it to be sunny, or see that it is sunny, the *intentional content* of my belief, desire, or perceptual state is that a certain state of affairs obtains. At first pass, a *mental representation* is an internal state of my mind/brain, which is characterized in functional or neural terms, and is a bearer of intentional content.[1]

Philosophical discussion of mental representations, and intentional content, has typically focused on two sorts of issues: “vertical” issues concerning the place of the mind in the natural order, and “horizontal” issues concerning the explanatory

[1] I will offer more thorough characterizations of the notions of intentional content and mental representation in Chapter 2.

role of mental representations in cognitive science (Rey, 2002). Let me elaborate on each of them.

Providing a *theory* of the intentional content of mental representations was one goal of the vertical project of "naturalizing" the mind: showing how our mental lives arise from physical matter. The *representational theory of mind* (RTM) is one prominent vertical approach. According to RTM, possession of mental states is to be analyzed in terms of realizing mental representations. For example, under RTM, all there is to believing some proposition P is for an individual to realize some mental representation that has the intentional content P, and plays the appropriate functional role in our psychology (Field, 1978; Fodor, 1987). If we accept RTM, then mental states reduce to mental representations, and the task of characterizing the intentionality of the mind turns into the task of providing a theory of intentional content: a non-circular analysis of intentionality which appeals exclusively to (non-intentional) properties of the natural world (Fodor, 1984). This project is not as popular as it once was (for criticisms of the project, see Stich, 1992; Tye, 1992), but it is still common to tie—or even pigeon hole—theories of content to the aims and ambitions of this project (see: Cohen, 2004; Rupert, 2008; Shea, 2013). In contrast, the horizontal project of determining whether or not mental representations are indeed indispensable to some of the explanations of cognitive science is, at present, comparatively fashionable (Burge, 2010; Orlandi, 2014; Ramsey, 2007). For while few would doubt that cognitive science is knee deep in internal states that are *called* "representations", or are described using intentional idioms, it is quite a separate question whether these posited states, in order to be explanatory, must have the distinguishing features of mental representations. In other words, while one can wonder whether (metaphysically) the mind is representational, one can also ask whether (explanatorily) the representational posits of cognitive science must be mental.

At root my project is horizontal, since it pertains to the explanatory role mental representations play in one corner of cognitive science. Nor is vision science a particularly isolated corner, since the explanatory practices of the field have often been treated as symptomatic of the larger role that mental representations do (or do not) play in cognitive science (e.g. Burge, 1986; Egan, 1995, 2010; Ramsey, 2007).[2] As stated earlier, my ultimate aim is to show that two common themes to the story of vision are in tension, and to suggest how the tension might be resolved. This aim is philosophically interesting for two reasons.

First, my defense of the information-processing narrative in vision science provides a counterexample against those who have argued, on very general grounds, that mental representations play no crucial role in the explanations of cognitive science (Chomsky, 1995, 2000; Egan, 1995, 2010; Ramsey, 2007). At least with respect to explaining facts about object vision, mental representations are indeed an indispensable posit (or so I will argue). Second, it has sometimes been claimed that theories of content, which have been developed for vertical endeavors, are simply irrelevant to horizontal ends (Burge, 2010; Stich, 1992; Tye, 1992). The reasons for this "perpendicular" view of theories of content tend to vary, but generally rest on the fact

[2] By "explanatory practices" I mean all the scientific work (theoretical, experimental) that goes into developing and testing different explanatory hypotheses.

that theories of content have typically been developed to address metaphysical, as opposed to explanatory, issues. Thus, they are simply beside the point when it comes to evaluating what sort of explanatory role mental representations may play in cognitive science. In contrast, my arguments provide a case-study in support of a "parallel" view, since theories of content provide strategies to relieve the tension I identify in the information-processing narrative of vision science. According to this parallel view, if certain explanations within cognitive science make assumptions about how content is fixed, then theories of content (and the challenges they face) might indeed be quite relevant to horizontal endeavors.

1.3 Outline of the Project

My project can be thought of as proceeding in three stages, which I will now outline, before providing brief summaries of the chapters to follow. The first two stages relate to showing why The Distal Object Thesis, and hence a notion of mental representation, is indispensable to explanations of object vision. The third stage shows why the Distal Object Thesis and Transformational Thesis are in tension, and how theories of content provide strategies for addressing the tension.

The first main conclusion that I defend pertains to general features of the information-processing framework in which research on object vision is embedded. Focusing primarily on the well-known work of Marr (1982), it has been vigorously debated whether the internal "representations" that the framework makes appeal to must be mental representations, or require any notion of representational content at all (Burge, 1986; Chomsky, 1995, 2000; Egan, 1995, 2010). I believe that the framework, as characterized by Marr, does minimally require a distinct content-component, and so makes an indispensable appeal to a notion of internal representation. However, whether these internal representations are mental or not will be wholly depend on the the visual phenomenon one is trying to explain. Thus the conclusion of the first stage of my argument is as follow:

> (1) A notion of internal representations is indispensable to the information-processing framework in vision science.

Conclusion 1 is enough to establish that some notion of representational content is indispensable to the information-processing framework (at least, according to Marr's influential characterization). But a further argument is needed for why a notion of mental representation—or more specifically, *perceptual* representation—is indispensable to explanations of object vision. Providing such an argument is the focus of the second stage of my project. A classic argument for why vision is both objective, and representational, relates to certain facts about the *explananda* of vision science. One of the most salient features of visual perception is its constancies: the fact that what we see remains unchanged, or invariant, under transformations of the sensory input. The traditional argument within perceptual psychology for objectivity in vision is that because what we see—the "object" of our perception—remains constant across these transformations of proximal stimulation, the object must therefore be something in the distal world (Brunswik, 1940; Cassirer, 1944; Thouless, 1931). Let us call this the *argument from constancy.* Recently, Burge

(2010) has revived this argument in making the case that perceptual representations are indispensable to the explanations of perceptual psychology, including vision science. While I am critical of Burge's attempted revival, I do believe that this sort of argumentative strategy can succeed when it comes to the viewpoint invariance exhibited by visual object recognition—our capacity to visually identify and categorize objects across changes in object orientation and viewing distance, retinal position, and surface illumination (DiCarlo et al., 2012). So an argument from *object* constancy can be made that supports the following conclusion:

> (2) A notion of perceptual representation plays an indispensable role in the explanation of visual object recognition.

My argument for Conclusion 2 provides a defense of the indispensability of the Distal Object Thesis to the explanation of some core aspects of object vision. In the third stage of my project, I show why the Transformational Thesis is in conflict with Conclusion 2 (and hence the Distal Object Thesis), and suggest how the conflict might be resolved.

According to informational theories, the content of a mental representation is fixed (at least in part) by a reliable causal relation between a state and what it represents (Dretske, 1981, 1988; Enç, 1982; Fodor, 1987, 1990). Loosely, the idea is that events carry information about what reliably causes them to occur. According to informational theories, whatever else the representing relation is, it is a kind of causal, information carrying relation between a representation and its content. For example, if I perceive a cat in front of me, and token my concept CAT, the reason that CAT is about cats, as opposed to something else, is because of the reliable causal relation between cats and CAT.[3] However, informational theories face a well-known difficulty, the *disjunction problem* (Fodor, 1984, 1987). A representation can be tokened, or activated, in a reliable manner by many causes, not all of which are part of the content of the representation. Raccoons or possums, under appropriate viewing conditions (e.g. in a back alley on a dark night) might reliably cause tokenings of CAT. Likewise, an appropriately well-placed knock to the head, when repeated, might similarly cause me to token the concept. It would seem that these causes, though reliable, are not content determining in the same way as the relation between CAT and cats. However if content is determined solely by a reliable, information carrying causal relation, then CAT represents the disjunction of its reliable causes.

I argue that the The Transformational Thesis presupposes an "crude" informational approach to representational content, according to which carrying information is indeed sufficient for fixing representational content. Since typical explanations of viewpoint invariance in recognition depend on the thesis, they also assume a "crude" informational theory of content. Therefore, these explanations run afoul of the disjunction problem. For this reason, absent a solution to the problem, they are explanatorily inadequate. Fortunately, existing information theories provide some strategies for addressing the disjunction problem, and are therefore of relevance to theories and models of object recognition aimed at explaining viewpoint invariance.

[3] By "concept" I mean categorical representations that are the vehicles of thought (Fodor, 1975). I use small caps to indicate when I am referring to the concept, as opposed to its content.

The conclusion of the third stage is as follows:

> (3) Informational theories of intentional content can contribute to the those explanations of visual object recognition for which the concept of perceptual representation is indispensable.

My argument for Conclusion 3 shows how informational theories might help resolve the tension between the Distal Object Thesis and Transformational Thesis. The breakdown of subsequent chapters is as follows:

In Chapter 2, I elaborate on my argumentative strategy for defending Conclusions 1 and 2, and lay out some different "recipes" for the notions of internal and perceptual representation. I then relate some potential anti-representationalist objections to my strategy and recipe. According to these objections, key ingredients from my recipe are never satisfied by posits within the explanatory framework of mainstream vision science (e.g. Chomsky, 1995; Egan, 1995, 2010; Ramsey, 2007).

In Chapter 3, I make my argument for Conclusion 1, focusing on the work of Marr (1982). I argue that Marr makes an indispensable appeal to a notion of internal representation. Central to my argument is showing that there is a content-component to Marr's approach. This is sometimes considered a "standard" interpretation of Marr's work, but I believe previous arguments have failed to identify the proper position of this component in Marr's multi-level approach to explanation. I also argue that the anti-representationalist objections reviewed in Chapter 2 do no justice to the information-processing framework, as characterized by Marr.

In Chapter 4, I take a critical look at arguments from constancy. While the argumentative strategy has been discussed for the better part of a century, I do not believe it has ever been articulated in careful detail. Hence in this chapter I offer an explicit (re)construction for the argument from constancy. I also identify some prima facie difficulties for this style of argument. Each of these challenges suggests a reason for rejecting an "objective" reading of perceptual constancies, which is required to ground arguments from reference. Having presented these challenges, I argue that the recent version of the argument defended by Burge (2010), while promising, cannot avoid them.

In Chapter 5, I make my argument for Conclusion 2. I show that when one focuses on the viewpoint invariance of object recognition, an argument from (object) constancy has promise. Once one incorporates research on object persistence and object-based attention in vision, and recognizes the role of perceptual learning in object recognition, my argument can meet the challenges from the previous chapter. Thus, in this chapter we see why a notion of perceptual representation (and hence the Distal Object Thesis) is indispensable to explaining one aspect of object vision—visual object recognition.

In Chapter 6, I make my argument for Conclusion 3. I spell out in more detail the nature of the disjunction problem and its significance for informational theories of content. I then argue that, due to the Transformational Thesis, research within the information-processing framework appears to presuppose a crude informational theory of content. For this reason, I believe explanations of the viewpoint invariance of object recognition require a solution to the disjunction problem. I then discuss different strategies for solving the disjunction problem; in particular, learning-based

(Dretske, 1981), teleological (Dretske, 1988; Neander, 1995, 2012), and counterfactual based solutions (Fodor, 1987, 1990). I evaluate these proposals based on how well they accord with facts about object recognition, and key elements of the story of vision.

So that is how I plan to proceed. So much for the threats and promises. Let's begin.

Chapter 2: Recipes for Representation

2.1 Introduction

As I outlined in Chapter 1, my project proceeds in three argumentative stages, each ending with one of my main conclusions. These conclusions are:

> (1) A notion of internal representation is indispensable to the information-processing framework in vision science.
>
> (2) A notion of perceptual representation is indispensable to explanations of visual object recognition.
>
> (3) Informational theories of intentional content can contribute to the those explanations of visual object recognition for which a notion of perceptual representation is indispensable.

In this chapter I do three things. First, I present and defend my argumentative strategy for Conclusions 1 and 2. Second, I spell out the notions of internal and perceptual representation that I appeal to in Conclusions 1 - 3 in a bit more detail. Third, I review some challenges that pose a threat to my arguments for Conclusions 1 and 2, and which I will need to overcome.

At the outset I should say that my discussion will not be exhaustive. I do not mean to provide a review of all the things cognitive scientists and philosophers have in mind when they talk about "representations". I only mean to elucidate the two notion that are relevant to Conclusions 1 - 3, and only in so much detail as suffices for making my arguments. What I propose to offer are some *constitutive ingredients* (Burge, 2010; Rey, 1997) for different notions of representation so that we can have some grounds for determining whether one or another notion is indispensable to some explanation in vision science. My approach is somewhat cumulative. There can be "representations" in more general and narrow senses. If we start with some base ingredients, many things in the world, both natural and man-made, qualify as representations. Add a few more ingredients, and one has what is needed for an internal representation in a brain like ours; a few more, a perceptual representation in the visual system. This is not the only way to try and get clear on the relationship between different notions of representation, but I think it is a reasonable approach, given my aims. So, in other words, here I offer a sampling of "recipes" for representation.

The rest of the chapter is structured as follows. In Subsection 2, I present my argumentative strategy for Conclusion 1 and 2, which turns on providing recipes of constitutive ingredients for different notions of representation. I also try to dispel some philosophical concerns one might have with my strategy (Mallon et al., 2009; Stich, 1996), and describe how it relates to the explanatory role of notions of representation in cognitive science. I further specify some "generic" ingredients required for my recipes. In Subsection 3, I describe in more detail the notion of *internal* representation I will be relying on in my argument for Conclusion 1. In Subsections 4 and 5, I enumerate my ingredients for *perceptual* representation. First, in Section 4, I offer up some "staple" ingredients for making a representational mental, which are inspired by some other recent recipes (Burge, 2010; Orlandi, 2014; Ramsey, 2007). Then, in Subsection 5, I provide some further "special" ingredients for perceptual content. I also distinguishing perceptual representations from mere sensory registers of proximal inputs. In Subsection 6, I review some anti-representationalist challenges (Chomsky, 1995, 2000; Egan, 1995, 2010; Ramsey, 2007), which if successful, would undermine Conclusions 1 and 2. In Subsection 7, I conclude the chapter.

2.2 How to Argue About Representations

My arguments for Conclusions 1 and 2 depend on carrying out a certain kind of procedure that can be broken down into four steps (cf. Cummins, 1989, p.145; Ramsey, 2007, p.10).

Step 1: Specify constitutive ingredients for a notion of representation.

Step 2: Identify what role some purported representation plays in some explanation (or framework) in cognitive science.

Step 3: Determine what ingredients something must have in order to play the explanatory role of the purported representation.

Step 4: Compare the constitutive ingredients with those that are necessary for playing the explanatory role.

If the ingredients that are necessary for playing the explanatory role are co-extensive with those that demarcate the target notion of representation, then we have an argument for why the relevant notion of representation is indispensable to the explanation (or framework) in question. My arguments for Conclusions 1 and 2 rest on carrying out this sort of procedure for notions of internal representation and perceptual representation (my argument for Conclusion 3 rests on a similar approach, but relating to theories of content).

A main objective of the present chapter is to provide some constitutive ingredients for different notions of representation, as required by Step 1. However before I settle into this task, I need to elaborate on and defend my argumentative strategy. In this section I first try to assuage some philosophical concerns one might have with my approach, and the very idea of trying to offer "recipes". Second, I described what sort of explanatory role I think representations tend to play, in general, in cognitive science. I also point to the sort of evidence one needs in order to show that certain ingredients are required for playing this role. Finally, I offer some assumptions about the general *types* of ingredients I think are needed when developing a recipe for representation.

2.2.1 Representational Recipes and Matters of Taste

My strategy bears a certain resemblance to a style of argument that is quite common in philosophy, and which has come under criticism as so-called "arguments from reference" (Mallon et al., 2009, Stich, 1996).[1] In general terms, arguments from reference can be thought of as a procedure that has three steps:

> **Step R1:** Assume (implicitly or explicitly) a substantive theory of reference for a crucial term or concept.
>
> **Step R2:** Argue that, given the theory from Step R1, the referent of the term or concept has certain properties (e.g., the term/concept does or does not refer).
>
> **Step R3:** Draw some metaphysical or epistemological conclusion (e.g., that the referent of the term or concept does/does not exist).

This sort of argument is especially common when philosophers want to draw strong ontological conclusions. The classic example of this strategy are arguments for eliminative materialism, the view that our commonsense ("folk") psychology is a deeply false theory, and that the states it posits, such as beliefs, desires, and emotions, do not exist (Churchland, 1981; Stich, 1983). Traditional defenses of the view amount to arguments from reference: they move from the assumption of a descriptivist theory of mental terms (e.g. for *belief*), and the fact that nothing posited by neuroscience or cognitive psychology satisfies the descriptions for these terms (i.e., the terms do not refer), to the conclusion that mental states do not exist (e.g., there is no such things as belief).[2] The problem with this argument

[1] In his review of Ramsey (2007), Sprevak (2011) also notices the structural similarity between the strategies.

[2] Briefly, under a descriptivist theory of reference, the meaning of a singular term (or "name") is specified by some description (definite or indefinite), and the term refers to whatever things satisfy

is that if one assumes a different theory of reference at Step R1, one can end up drawing a different ontological conclusion at Step R3 from the very same facts at Step R2 (Stich, 1996). Other examples of arguments from reference can be found in philosophical debates about scientific realism, moral realism, and the ontology of race (Mallon et al., 2009).

While there is a definite similarity between my strategy and some arguments from reference, one crucial difference is that I do not intend to draw any ontological conclusions. Even if some notion of representation (internal, perceptual or otherwise) was entirely dispensable to cognitive science, we would not be obligated to draw the ontological conclusion that the species of representation does not exist. Thus my argumentative strategy is closer to those aimed at determining whether certain notions, such as "emotion" (Griffiths, 1997, 2004), or "concept" (Machery, 2005; Machery et al., 2009), are indispensable explanatory kinds for cognitive science. Whatever the merits of these projects, they are silent about matters of ontology.

However, even if I eschew the dubious third step of arguments of reference, some of the criticisms of the strategy are nonetheless relevant to evaluating my approach. I will go through three of them that seem to me especially pertinent, as they relate to what sort of ingredients I can or should offer in my recipes. In each case, I will argue, the issue largely boils down to matters of taste. And whatever one's preference, I think my approach is palatable.

the description.

Issue 1: Referential Assumptions. The first issue relates to what sort of assumptions I might be making regarding theories of reference. While the problem is typically related to the reference of singular terms, the same sort of considerations apply with respect to concepts. By offering recipes for notions of representation I seem to be assuming something roughly descriptivist: representational concepts relate to a description that specifies a set of properties (i.e., ingredients), and something is within the extension of the concept if and only if it satisfies the relevant description. Even though I am not drawing any ontological conclusions, one might still worry that I risk predetermining my conclusion due to my referential assumptions. I have a number of replies to this worry.

Although I am offering constitutive ingredients for certain notions, I do not think I need to make any firm commitments regarding the correct theory of reference. I am happy to adopt the view that I am *stipulating* a descriptivist approach. Furthermore, precedent suggests that such an approach usually works in the favor of those who aim to draw a negative conclusion: that some term does not refer (Churchland, 1981; Stich, 1983), or some kind is not explanatory (Griffiths, 1997, 2004; Machery, 2005; Machery et al., 2009). Indeed, this seems to be the case with some recent arguments against the utility of notions of representation in cognitive science more generally (Ramsey, 2007), and in vision science in particular (Orlandi, 2014). The reason for this negativity is that laying out a descriptive recipe sets a rather high bar—it is doubtful that all the ingredients are ever in fact satisfied. So I do not think I am doing myself any favors by assuming a descriptive approach, beyond the virtue of being explicit about my assumptions.

Still, if a quasi-descriptivist assumption is considered to unpalatable for some, I believe my arguments can still be run if one assumed a different approach to concept application, such as a causal-historical approach.[3] For even under such a view, a description might be used initially to fix the referent of a term or concept, although the ingredients that make up the description are not constitutive. Under such a view, I believe that a recipe would still have *discriminative* utility, for distinguishing between cases. Of course my conclusions would be weaker (satisfying the description merely gives us good reason to think a notion is indispensable), but I think these weakened conclusions would still be philosophically interesting.

Issue 2: Folk Psychological Commitments. Granting a vaguely descriptivist approach to Step 1, a second issue relates to what sort of facts are supposed to ground the description. Specifically, with respect to mental representation (and hence perceptual representations), the natural starting point are intuitions drawn from folk psychology. There are two related concerns one may have with relying on folk judgment in the development of some of my recipes.

First, one might argue that if my goal is to articulate any notion of "representation" utilized in cognitive science, our folk understanding will be off little help, since our interest will be in what cognitive science tells us representations "really

[3] Briefly, causal-historical approaches hold that a singular term is introduced by a speaker to refer to something. Subsequent users of the expression pick out the same referent due to a causal chain that leads back to the speaker who legislated the referent. And although a description might be used to initially fix the referent, it is the thing itself that is referred to, and not whatever happens to satisfy the description (Kripke, 1980).

are", and not what folk psychology suggests they might be (Stich, 1992, p.252). This is well and good when it comes to a notion of internal representation (as we shall see below), but the problem is that without some connection to our folk understanding of mentality, it is not clear what basis we have for talking about *mental* representations—including perceptual ones (Ramsey, 2007). I think a reasonable view is that our folk judgments at least provide a starting point for constitutive analysis, until we have good reason to reject them (Rey, 1997, p.34). This sort of provisional acceptance of everyday theory and observation can also be found in vision science. For example, Marr (1982, p.331-332) recognized that everyday experience often provides the starting point for investigating how we see.

Granting that it is reasonable to appeal to folk judgment (to some degree), a second concern is that one may be hard pressed to find *agreement* with respect to our judgments about the mind (Stich, 1996). Thus one might despair that I am no more likely to succeed with respect to a notion of perceptual representation. While I agree that intuitive judgments may often diverge, I think this just provides motivation for me to be be clear about my own assumptions. It is for this very reason that going forward I will be as explicit as possible about the notion of perceptual representation I am relying on. I fully accept that there might be a plurality of notions available. Indeed, many distinct (but similar) recipes for perceptual representation can be found in the literature (see e.g.,Burge, 2010; Orlandi, 2014). Minimally my claim is that I have successfully isolated one of them, which (as I shall argue) seems to play an indispensable role in the explanation of object vision.

Issue 3: The Scope of the Ingredients. A final worry concerns just how elaborate I intend my constitutive analyses to be. Do I intend to offer an account of the essence of what it is for something to be a representation? Do I promise necessary and sufficient conditions? A careful conceptual analysis? A definitive definition?

It should be clear by now that my ambitions are a bit more modest. I believe that it is enough that I offer recipes that allow me to correctly identify the clear cases of representation (of one sort or another) from the vague ones (cf. Rey, 1997, p.32; Quine, 1960). So, operationally, I am happy to treat my recipes as providing necessary and sufficient conditions which discriminate between the cases of interest. Consider by way of comparison, one recipe for the Rickey, a cocktail first created in Washington, DC, in the late 19th Century:

- 2oz bourbon or gin.
- Half lime squeezed and dropped in the glass.
- Add ice and stir.
- Fill class with soda water.

A simple recipe, but there are borderline cases. Once while in DC I had a drink that included (amongst other ingredients) rye, chartreuse, fennel orange marmalade, and a blood orange IPA. Delicious, but I confess to being unsure (even after repeated sampling) whether it was indeed a Rickey, advertisement to the contrary. However, whether this particular concoction was a Rickey, it seems we can still clearly distinguish the Rickey from its cousin the Mojito (which includes rum, simple syrup, and muddled mint), and also recognize that some elements of a typical preparations are inessential (e.g., that it be served in a highball glass). The upshot of this example is that for the cases I am interested in, the recipes I will offer will suffice. For other borderline cases, I leave it up to (theoretical) taste.

2.2.2 Explanatory Roles and Evidence for Indispensability

Granting, then, that it is reasonable (given my strategy) to offer recipes for the notions of representation I intend to employ, one might next ask what sort of explanatory role I have in mind with respect to Step 2, or what sort of evidence would suffice for showing, at Step 3, that certain ingredients are required for a posited representation to play the role. Both of these issues are especially important when it comes to defending Conclusion 2, and showing that a notion of perceptual representation is indispensable to explanations of object recognition.

Regarding role, in many explanations in cognitive science representations (of whatever sort) appear as both *explanantia* and *explananda*—as explanatory posits and phenomena to be explained. On the one hand, representations are posited to explain certain psychological phenomena (perhaps as revealed by careful psychophysical experiments). And in order to explain these facts, the representations must have certain properties related to the phenomenon of interest. On the other, for representations that have been posited to explain the psychological phenomena, theories and models are offered of how the representations are structured, or are able to represent what they represent. So while the goal is to explain apparent facts about these representations they themselves are also (originally) theoretical posits. And by offering theories of models of how they work, we are simply fleshing out the explanation of the relevant psychological phenomenon that they were initially posited to explain (Ramsey, 2007, p.36).

This sort of "dual role" is not unique to the explanatory posits of cognitive science. Consider by way of comparison the notion of ion channel, which is a common example in the philosophy of science (Machamer et al., 2000). In very general terms, ion channels were posited to explain changes in the concentration gradients of different ions (inside and outside the cell body), which drive the action potential in neurons. For instance, depolarization of a neuron occurs when sodium ion channels open allowing sodium (Na^+) to flood the cell body. So sodium ion channels were a crucial posit in the explanation of the action potential. But one might also want to know *how* they carry out their functional role, and offer a theory of the structure of sodium ion channels, and how they work to gate ion flow. Thus, in a fashion similar to representations, ion channels were both posited as part of an explanation (of the action potential), but have also became a phenomenon to be explained (i.e., how they are structured).[4]

Representations also do double-duty in vision science. Research within the information-processing framework posits internal representations to explain facts

[4] There are of course some disanalogies. Psychological phenomenon have at least some connection to folk understanding, while ions are a phenomenon discovered by science. Also, arguably the ion channel example might involve roles at different levels (Craver, 2007), which may not be true of the two explanatory roles of representations in cognitive science.

about vision (a claim I defend in Ch.3). To explain particular visual phenomena, representations are posited with properties relate to the facts they are supposed to help explain. As we shall see later on, to explain facts about object recognition, representations of object identity and category membership are posited to explain recognition phenomena (e.g., why recognition performance is so invariant across transformations of stimulus viewpoint). In turn, theories and models are offered of how these object representations might be structured, such that they can play the role they are hypothesized to play.

In fact, the Distal Object Thesis and the Transformational Thesis each relate to one of the two explanatory roles of representations in vision science. First, the Distal Object Thesis points to what kind of representation must be posited to explain some aspect of vision. Second, the Transformational Thesis relates to what sort of explanations we are to give in light of the functional organization of the visual system. To presage later discussion, that these two theses relate to these different roles is important to the tension I will reveal between them.

But before I reveal this tension, I need to show that explaining object recognition requires positing *perceptual* representations, in a sense I will be articulating shortly. And prior to giving some substance to the notion, I think it is worth asking what sort of evidence would show that a notion of representation is indeed indispensable to an explanation. As we shall see in due course, it will be fairly easy to show that representations of the right sort are appealed to by explanations of object recognition; that is, that the explanations include the positing of what appear to be perceptual representations. But simply showing that a notion features in an explanation is not enough to illustrate its indispensability.

By way of comparison, here is what I take to be a rather bad way to argue if one hopes to show that some notion of representation is indispensable to an explanation in cognitive science: first, argue that representations of the relevant sort are posited in some successful or fruitful explanation of a psychological phenomenon; and next, conclude that the notion of representation is therefore indispensable to the explanation. This argument is bad because the mere fact that a notion of representation features in an explanation does not show that it is indispensable. As Ramsey (2007, p.1) points out, it could be as unnecessary to the explanation as the notion of celestial sphere was to Copernicus' astronomy.

What will be needed instead is an argument for why having the ingredients for perceptual representation (which I present below) are *essential* to posited representations playing their role within an explanation. Of course, scientists also worry about what they "need" to posit to explain a phenomenon. Thus one might wonder whether, in arguing for the indispensability of perceptual representation, I will simply be defending my *own* view about how to explain object recognition. In which case, I am not commenting on the explanatory practices of vision science, but rather *engaging* in them myself.

I have two responses to this worry. First, I think one can distinguish between two enterprises that involve evaluating what sorts of posits are necessary for an explanation to be successful. One involves simply positing certain entities which one thinks are necessary for the explanation of the phenomenon of interest; another is to determine, in general, if certain posits are in fact needed for *any* explanation of the phenomenon one might hope to offer. So in the first enterprise one does both the positing and the explaining. In the second, one does neither.

I agree it would be problematic if I was engaged in the first endeavor, which I am not (as I hope will become clear in due course). Rather, I take myself to be engaged in the second, which I admit could also be considered as falling within the scope of vision science, in so far as it might relate to foundational questions about how the science tries to explain what and how we see. But it is also within the scope of philosophy, in so far as one might be interested in critically evaluating the sorts of explanations found in one branch of science. For example, a we shall see in the next chapter, Marr (1982) was clearly engaged in both enterprises: of offering both explanations of particular visual phenomena, and an approach for *how* we should explain visual phenomena. And this is in part why his work is a touchstone for both vision scientists and philosophers of cognitive science.

My second response is that it is not always terribly clear why vision scientists insist on positing "representations", in which case a fair amount of reconstruction might be required to see why a particular notion of representation is indeed indispensable. In particular, I will engage in this sort of extrapolation in my discussions of perceptual constancy in Chapter 4, and object recognition in Chapter 5. The theoretical connections I make could be exploited by those interested in offering their own explanations of facts about these visual phenomena, but that would be to adapt my arguments for a different enterprise than my own. Indeed, one could do the same with my recipes for representation, which I will now begin to develop.

2.2.3 Some Generic Ingredients

To get started on the right foot, I want to first offer a sort of recipe for what I will term a *generic* notion of representation.[5] This is intended to be a very general notion that specifies what *types* of ingredients my recipes for internal and perceptual representation must have.

One might think that in the broadest terms a representation is anything that has content, or a "semantics". After all a representation, whatever it is, cannot be what it is without being "about" something.[6] This is true whether we talk about street signs, entries in the data structure of a digital computer, or internal states of a brain or nervous system. Of course how one part of the world manages to be about another, and what kinds of things it represents, might vary a great deal between cases. But that some object or event in the world serves as a vehicle for content is surely necessary ingredient for a very general notion of representation. However I do not think it should be taken as sufficient (Fodor, 1990; Ramsey, 2007). Representations do not exist in a vacuum. They always represent to, or for, something.

By way of example, consider a magnetic compass, which is a plausible case of a representational device [7]. On the one hand, what the needle of the compass represents, its content, is the direction of the nearest magnetic pole—North, when in the Northern hemisphere. On the other hand, the representational function of

[5] Here I take inspiration from the notion of "generic computation" offered by Piccinini and Scarantino (2010), which I discuss in Chapter 3.

[6] Though see the discussion of "ersatz" representation below.

[7] I owe the example of a compass as a representational device to Ramsey (2007, p.29).

the device is to track geographic North.[8] The fact that the needle has some content is not sufficient for making it a representation. Suppose a child rubs a needle on a magnet, and places it on a floating leaf. The resulting magnetism in the needle will cause its tip to orient toward magnetic North. When constructed by the child, the needle and leaf is not a compass. But in a survival situation one might use the same trick as the child, and the needle would indeed function as a compass.

The importance of the compass example is that it illustrates that at least for some kinds of representation content is not enough. One must also show that the object or event has a *representational function*. This dual requirement seems to be generally true of the sorts of representations that interest us, including ones that are mental. For example, Dretske (1988, p.80) makes the point in terms of reasons:

> The fact that [reasons] have content, the fact that they have a *semantic* character, must be relevant to the kind of effects they produce.

And Ramsey (2007, p.27) makes the same point in more general terms:

> [T]o be a representation, a state or structure must not only have content, but it must also be the case that this content is in some way pertinent to how it is used.

So a recipe for representation requires two types of ingredients, relating either to content or representational function. And anything that has at least these two (types) of ingredients will qualify as a representation in a generic sense:

[8] We will ignore here the fact that what it tracks, and what it is used to represent, seem to come apart. See discussion below.

> *Generic representation:* an object or event that: (i) is a vehicle for content; (ii) plays some functional role in virtue of the properties of its content.

This very general notion does not stipulate what the content is, how it is fixed, what sort of function the object or event carries out, or the kinds of causal powers it possesses. Objects like street signs are subsumed, as are events like when the light changes in a stoplight. It also subsumes minds, computers, and other things that have internal state event types that (as we shall see) are considered representations.

In this and subsequent recipes, I do not intend to take a firm stand on exactly how we are to understand what functions are. For simplicity, I will assume that functions apply to elements of a larger structure or system, and that they are capacities of these elements that are subsumed under some sort of nomic generalization (Cummins, 1975, 1983).[9] I think that representational functions, whatever they are, are functions in this general sense. Street signs have their function in terms of the role they play in a larger system for regulating traffic; a compass needle has its function by being part of a system that is composed of, among other things, a windrose.[10] The only thing I have added to differentiate something as a *representational* function is that the capacity is connected to the content of the object or

[9] There are multiple forms of functional analysis that might be relevant, given my interest in the explanation of psychological phenomena (Piccinini and Craver, 2011). The notion of functional analysis I have described is a form of "task-analysis". I believe that one could rely on this or another version of functional analysis (or incorporate it into a mechanistic approach to explanation; Craver, 2001), with little influence on my arguments to follow.

[10] A "windrose" on a compass is the graphic that indicates the cardinal directions.

event—it has its functional role in a system in virtue of the properties of its content.

Having made the argument that representational function is an ineliminable ingredient for any notion of representation, let me differentiate my position from three other views I do not intend to endorse. Each of these pertains to the relationship between content and function in a representation.

*1. Teleology.*There is a line of thinking according to which a representation must have an appropriate function, but in different sense than I intend. According to this approach, the function of a representation is its teleological "proper function", and its content is "consumer-based" in that what something represents is determined by how it is used by the system in which the object or event is embedded (Millikan, 1984, 1989; Papineau, 1987). While usually presented as a theory of intentional content, this "teleosemantic" approach can also be considered as a very general recipe for representation (Millikan, 1984; Ramsey, 2007). The core of the view is that something has its (teleological) function and content in virtue of how it is selected or used, rather than what it does. Whatever the merits of this perspective, it is different than what I have in mind with respect to both representational function and content, since I am not relying on a notion of teleological function.

2. Conceptual-Role. According to conceptual-role theories, content is determined by is functional role(s) in a larger system (Block, 1986; Harman, 1982). I think that this sort of "content" that is determined by how a representation is used is different from what I have in mind when talking of content. Let me again illustrate using the example of a compass.

Like most representational devices compasses are designed and constructed with a particular functional use in mind, namely, to aid navigation; that is, they are constructed to have a representational function, of indicating the direction of the geographic poles, so that they can be *used* in navigate. However I can readily use a compass in other ways that take advantage of its representational function, but which do not involve navigation. For example, if you are building a house, and want a North-South exposure, then you can use a compass to orient your construction plan.[11] Likewise, a Muslim following the Five Pillars of Islam could also use a compass to find *Qibla* (the direction of Mecca) during prayer. Indeed, in general, if I need any information regarding direction that can be derived from the direction of magnetic North, I can use a compass, even if I do not plan on navigating.

The relevance of this distinction between role and use is that it suggests that we can differentiate between two candidates for the content of a representation (Cummins, 1996). On the one hand there is what I am calling the representational content of the compass, which is related to the cardinal directions, and on the other there is the particular "target" I might try to isolate by using it. This might be the direction I need to walk, align my building plan, or orient my prayer mat. How a representation is used determines the target, but not the content, of the representation, under my view. So far I have highlighted that many targets can go with one content, but sometimes the two simply diverge. For example, a magnetic compass is designed to indicate the nearest magnetic pole (assuming minimal magnetic dis-

[11] In fact, the magnetic compass was first invented for just such a function as part of *feng shui* geomancy in ancient China. It was centuries later the magnetic compass was combined with the windrose to create what we now consider the modern magnetic compass.

turbances nearby), and for use in navigation when trying to determine the direction of a geographic pole. But when near the Equator, or one of the Poles, it is of little value for determining the cardinal directions no matter the intended use. The point of all of this is that when I speak of content I do not mean a target, whereas under a conceptual-role approach to representations the two are seemingly the same.

3. Causal Powers. I have said that a representation should have its function in virtue of its content. Plausibly, the causal powers of something are related to its functional role. This raises the question of whether I think the content of a representation is somehow causally efficacious. Some have vigorously rejected this idea. Instead, the causal powers of a representation are exhausted by its physical or formal properties—at least when it comes to internal or mental representations (Dretske, 1988) or representations that are also considered to be computational symbols of some kind (Fodor, 1980). Others have argued that content is indeed causally relevant (Peacocke, 1994; Rescorla, 2014). While an interesting issue I intend to remain agnostic on the matter, as I believe either of the two alternatives would be consistent with my arguments to follow.

In summary, the notion of representation I describe below are all generic: their recipes include ingredients for both for content and representational function. But my generic notion should be distinguished from teleological, and conceptual-role based generic notions. And I am not taking any stand as to whether my generic notion requires that the content of a representation is causally efficacious. We now turn to some recipes for internal representation drawn from cognitive science.

2.3 Some Common Varieties of Representation in Cognitive Science

Notions of representation in cognitive science are legion. For example, when a linguist talks about "representations" of the syntactic structure of natural language, and a neuroscientist talks about what the firing pattern of a single-cell "represents", it is not clear that they have the same notion in mind. Nonetheless, I think it is possible to articulate some very general notions, which crucially are not motivated by any considerations regarding mentality or folk psychology. In particular, in this section I provide a recipe for a fairly broad notion of *internal* representation, which I contrast with a different sense of "ersatz" representation that is sometimes found in the literature.

2.3.1 Internal Representation

As we have already seen, I think recipes for representation need to have ingredients relating to both content and function. In cognitive science it is common to talk of "internal representations" of various kinds. At least historically, such talk has been the result of taking the mind or brain to be a computer of some kind, in which instances of internal representations are simply computational symbols that are implemented by the brain (Newell, 1980; Pylyshyn, 1984). While this historical precedent is important, I would like to rely on a more minimal conception of internal representation that does not make overt reference to computation. In what follows, I will rely on the following recipe for internal representation:

> *Internal Representation:* any internal state type of a system; that (i) is a vehicle for *original* content; and (ii) performs a "stand-in" function with respect to the internal operations of the system.

Let me unpack this analysis. By "system" I mean something quite broad. It can be characterized in either concrete (a nervous system) or more abstract terms (a multiply realizable functional architecture). I do not assume any amount of organizational complexity. Bacteria or brains may do. The internal state type is an event type that can be tokened; that is, it is something that a component or part of the system enters into, based on internal or external causal factors.

The first constitutive ingredient I have identified is that the content be "original". Content is considered *derived* when it is a product of the artifice or convention of an agent or group of agents; that is, it depends on the representational (and indeed, intentional) capacities of an agent. In this broad sense, various representational objects or devices have content that is derived: street signs, stoplights, thermostats, compasses, and even digital computers. In contrast here I presume that the content in question is not derived but *original*: what a particular internal representation is about is not assigned by artifice or convention of an agent, or set of agents. Thus, however the content is fixed, it is by a different means than things like road signs, which have derived content resulting from human convention.

One might find it surprising that I associated original content with a notion of internal representation. Typically originality is identified as a distinguishing ingredient of the intentional content of mental representations. For example, Ramsey

(2007, p.17) states that:

> …the aboutness of a word or a road sign is thought to exist only through the aboutness of our thoughts—in particular, the aboutness of the thought that these physical shapes stand for something else…Only thoughts and other mental representations are assumed to have what is called "original" or "intrinsic" intentionality.

However I think we should separate the question of whether the content of some representation is original (or intrinsic) from the question of whether it is intentional. It is quite reasonable to suppose that there are states of physical systems that might be representations in a generic sense, but do not have derived content. For example, perhaps individual neurons are a simple kind of representational device, which are senders and receivers of information about action potentials (Cao, 2012), or even the most primitive unit for perception (Barlow, 1972). In such a case, what the neurons represent is not dependent on the goals, thoughts, or intentions of an agent, yet neither is the content that they encode necessarily intentional. Along similar lines, the "symbol grounding problem" (Harnad, 1990) in artificial intelligence concerns how an artificial agent can autonomously determine the contents of their own internal states—in other words, have original content (Taddeo and Floridi, 2005). In this case, there need be no presumption that we are speaking of mental representation, and artificial minds.

So it seems that the notion of original content and intentional content can, at least conceptually, be teased apart. Hence, I identify original content as an ingredient for internal representation, and recommend identifying some other constitutive ingredients for intentional content, as we shall see below.

The second constitutive ingredient relates to the functional role of internal representations. A very general representational function is that some event or object serves to "stand-in" for something else. This notion of stand-in function is regularly invoked in cognition science, though it is not necessarily easy to characterize (Bechtel, 1998). Newell (1980, p.156) offers one influential characterizing of the function, which he calls "designation":

> An entity X designates an entity Y relative to a process P, if, when P takes X as input, its behavior depends on Y.

The core of this description is that for the process P, X takes the place of the entity Y with respect to the process. Many representations might have this function, especially if we presume X does not need to obtain solely with respect to entities of some kind (e.g., if it could also be a property or state of affairs). Newell was motivated by an analogy to symbols in a digital computer, but one can get at the same idea more intuitively: something has a stand-in function when we directly reason about it to draw conclusions about something else. In this respect, something functions as a "surrogate" that allows us to:

> ...reason directly about a representation in order to draw conclusions about the things it represents...In such cases we use one sort of thing as a surrogate in our thinking about another...(Swoyer, 1991, p.449).

For example, arguably a compass functions as a stand-in (or "surrogate") for the direction of geographic North when I use it to navigate while hiking. In Newell's terms, the needle of the compass (X) has the function to designate the direction of the nearest magnetic pole (Y), with respect to my deliberation process (P) about how to modify my route. Or in more intentional terms, by reasoning about the needle direction, and what it tells me about the polar directions based on my current position on the trail, and comparing this information to landmarks and polar coordinates topography of the trail map, I can draw a conclusion about how to make my way back to the car.

Hopefully these gestures are sufficient for grasping the idea of a "stand-in" function. So characterized, one might wonder if a "stand-in" function is *the* way to think about representational functions in general. For example, Haugeland (1991, p.62) states that something which: "stands in for something else in this way is a *representation*; that which it stands in for is its *content*; and its standing in for that content is *representing* it." So perhaps being a stand-in is indeed the only function any representation ever has. I am non-committal on this issue. And rather than attempting a more complete characterization of the ingredients for internal representation, let me illustrate the notion using two kinds of internal representation often posited in cognitive science.

The first kind of representation is of an internal state of a system that is part of a stage in the transformation of information. This internal state functions as a stand-in for something, as part of the functional organization of the larger system in which it is embedded. For this reason, it is common to talk of this sort of representation as "encoding" certain information about a stimulus, which is then "decoded" by a later stage of processing, or as either the input or output representation of a process. This notion of what we may call, following Ramsey (2007), an "io-representation" is intimately connected with the information-processing framework in vision science (Marr, 1982), as we shall see in the next chapter. The second kind of representation involves modeling or simulating some domain. This kind of representation rests on some form of structural mapping, or isomorphism between properties of the representation and the domain it represents (Gallistel and King, 2009; Palmer, 1978; Shepard, 1984). Following Ramsey (2007), we may call these "s-representations". Although the kind of internal representation we are interested in will be io-representations, s-representations also qualify as a form of internal representation as I have defined them.

These two notions of internal representation are sometimes identified with specific commitments about the functional architecture of the brain (e.g., Ramsey, 2007). As a matter of history it is true that the notion of internal representation is deeply connected to the view that the mind/brain is (in some sense) a computational system (Fodor, 1980; Newell, 1980; Pylyshyn, 1984).[12] In this respect the notion of internal representation I have presented is closely connected to the idea of a computational symbol over which the rules and operations of a system (i.e., the

[12] This statement is ambiguous, since I am fudging a distinction between computational functionalism as a metaphysical thesis about the mind, and computationalism as a view about explanation (Fodor, 2000; Piccinini, 2010). However the difference between these views does not matter for present purposes.

computations it carries out) are defined. However I want to resist presupposing a notion of computational symbol when talking of internal representation. Perhaps the internal representations I am describing, and the architecture they require, qualify as computational in a generic sense (Piccinini and Scarantino, 2010): they can be considered medium-independent vehicles that are manipulated as part of some process that has the function of manipulating them in accordance with rules defined over the vehicles. But I do not want to presuppose more than this.

What I do think is important is that the notion of internal representation I will be employing does not have its origin in folk psychology, but rather emerged in cognitive science from reflection on the organization of the sorts of (computational) devices we commonly engineer for detecting and representing signals, and storing and analyzing data.

2.3.2 Ersatz Representation

Having specified a notion of internal representation that was gleaned from cognitive science, I want to contrast it with a different notion that is also sometimes encountered in the literature. According to this notion, "representation" is just a synonym for a "computational symbol", which need not have any sort of content or "semantics". I will call this the "ersatz" notion of representation, since the internal states of systems that it picks out are not representations in the minimal generic sense I spelled out earlier (because they lack content).

It is an ongoing debate in the philosophy of computation whether you can have computation without representation (Piccinini, 2008). Matters depend, of course, on the precise way in which each notion is characterized. The relevance of the ersatz notion for present purposes is that some have argued that it is the *only* notion that is in fact employed—or needs to be employed—in the explanations of cognitive science (Chomsky, 1995, 2000; Stich, 1983). I am happy to recognize this very general notion of internal "representation" whereby one means any internal state of a system over which some sort of (computational) operation is defined, but such a state would not be a representation of the sort I have in mind.

So I mention this usage of the term 'representation' only in order to acknowledge it, and set it aside.

2.4 A Recipe for Perceptual Representation I: Some Mental Staples

With a notion of internal representation at our disposal, the next step is to enumerate the further ingredients that are required to make a perceptual representation. I have broken this task into two parts: (I) providing some "staple" ingredients for when an internal representation is mental; and (II) providing further "special" ingredients that distinguish when an internal representation of a sensory system is not just mental, but also has perceptual content. I take up (I) in this section, and (II) in the next one. And although my aim in this section is to provide some of the constitutive mental ingredients for perceptual representation, I will proceed by in part considering ingredients for mental representation more generally.

At root a mental representation is an internal representation (characterized in functional or neural terms) that is a vehicle for intentional content. This makes sense, since intentional content, or intentionality, is typically considered a mark of the mental. It is also common to distinguish the sort of intentionality in question as being original, which we have already included in the recipe for internal representation. So other ingredients must be identified to distinguish intentional from merely original content. Before doing so, let me make two clarifications regarding the character and scope of the ingredients I will offer for mental representation.

First, in providing a set of mental ingredients for representation, I am neither presupposing a commitment to the representational theory of mind (RTM), nor attempting to offer a set of ingredients for naturalizing intentionality. As my aims are horizontal, the recipe I offer is not intended as a constitutive analysis of mental states, or as account of how intentional content can be fit into the natural order. In other words, the ingredients I offer are for determining whether internal representations posited in vision science are mental, not for whether mental states posited by folk psychology are representational.

Second, the sorts of ingredients I offer are for a restricted class of mental representations, which subsumes perceptual representations. The first restriction is that a mental representation must be directed at the external world, or have a “mind-to-world” direction of fit. This restriction excludes mental representations that have a “world-to-mind” direction of fit, like those we might connect to desires, motivations, intentions, goals, and motor planning and action. The second restriction is that the mental representation have some external referent. This restriction is necessary because the Distal Object Thesis presumes some distal referent.[13] In fact, going forward, I will usually take "content" to simply specify the external referent of a representation. Given this restriction, my recipe likely excludes many kinds of mental representations, due to what they are about. For example, we think of love, justice, and logic, with perhaps equal clarity and vigor, as we do fictional characters such as Sherlock Holmes (or to use a more current example, Harry Potter) or non-existent things such as the planet Vulcan. Representations that underlie these capacities may be perfectly mental, but they are not the sort of cases I want to capture with my recipe because they lack a distal referent.

When operating within these restrictions, there is something of a consensus as to the sorts of ingredients that are required for intentional content and mental representation, whether one is talking about perception (Burge, 2010; Orlandi, 2014), or cognition (Ramsey, 2007), with similar ideas being expressed earlier by Fodor (1986). The three "staples" I will focus on are *objectivity*, *robustness*, and *correctness* conditions. Varieties of each provide important constitutive (mental) ingredients for perceptual representation.

2.4.1 Objectivity

The first constitutive ingredient is related to the Distal Object Thesis: the internal representation must be about some "objective" property instantiated in the world. Described in this way, I take objectivity to be a property of the intentional

[13] The thesis, recall, states that: "the end-stage internal representations that underlie object vision represent properties of entities in the distal world."

content of some mental representations.

Following Burge (2010, p.47) there are two notions of objectivity that are of potential relevance. First, we might think of a subject matter as being objective in the sense that it is *mind-independent*. Read narrowly, this notion precludes anything that results from human artifice as being objective. Second, we might think of a subject matter as being objective in the sense that it is constitutively *non-perspectival*. The latter notion is broader, and is the one I will rely on. Nagel (1980, p.77) provides a useful description of this sense of objectivity:

> To acquire a more objective understanding of some aspect of the world, we step back from our view of it and form a new conception which has that view and its relation to the world as its object. In other words, we place ourselves in the world that is to be understood. The old view then comes to be regarded as an appearance, more subjective than the new view, and correctable or confirmable by reference to it. The process can be repeated, yielding a still more objective conception.

Under this conception of objectivity, while minds might themselves be perspectival, what they represent is not. Also, under this conception, objectivity also comes in degrees. If we think of perspectives as pertaining to how we perceive, then multi-modal internal representations that integrate information from multiple sense modalities will be more objective than unimodal ones. In turn amodal internal representations, such as concepts, are more objective still. For the subject matter of what we see, or hear will likely be dependent on facts about how our sensory systems work in a manner that more abstract categorical representations are not.

The content of an internal representation will be what I will call *perceptually* objective when it is non-perspectival with respect to a sensory system. For the visual system, this means the representation transcends viewpoint, in that the content is independent of particular vantage points by which something might appear to us: the projected retinal size and position, its particular orientation in the picture- and depth-planes, and its apparent illumination and distance from the viewer. In other words perceptually objective content is maximally non-perspectival with respect to the dimensions of perspective within the visual system. Going forward I will take perceptual objectivity to be a constitutive mental ingredient for the intentional content of perceptual representations.

Just because I think the representations of interest are about non-perspectival properties or states of affairs does not mean I think they must be "veridical" in the following sense: what we represent is what we *think* we represent. Sometimes the very idea of mental representation is tied to the idea that our intuitions about what we represent must by and large be correct (Akins, 1996). For example, one might claim that when we perceive color, we cannot be mistaken about what we think it is we perceive (Mendelovici, 2013). I think this view is itself mistaken. To revisit our earlier discussion of theories of reference, it could be that many of our explicit beliefs about what we represent are wrong, even if what we represent is indeed some objective property of the world. So veridicality in this sense, I think, is not an important mental ingredient for the sort of perceptual representations I have in mind.

2.4.2 Robustness

The second ingredient I will emphasize relates to the fact that what a mental representation is about can, in some sense, be "absent" (Orlandi, 2014; Pylyshyn, 1984). The connection between absence and intentionality was recognize by Brentano (1874) who is largely credited with reintroducing the notion of intentionality to philosophical scrutiny. Brentano pointed out that the objects of paradigmatic propositional attitudes, such as belief and desire, have an "intentional inexistence". What exactly Brentano meant by this phrase is not entirely clear (for some discussion see: Crane, 1998; Rey, 2012). Plausibly, part of what he had in mind was that the objects of our attitudes need not exist, as captured vividly by Chisholm (1952, p.56-57):

> [the propositional] attitudes can truly be said to have objects even though the objects which they can be said to have do not exist. Even if there weren't any honest men, for example, it would be quite possible for Diogenes to *look* for one. Diogenes' quest has an object, namely an honest man, but, on our supposition, there aren't any honest men. . . But mere physical phenomena, on the other hand, cannot thus "intentionally contain an object in themselves." In order for Diogenes to sit in his tub, for example, there must be a tub for him to sit in.[14]

[14] Diogenes of Sinope (412 - 323 BCE) was one of the founders of the cynic school of philosophy in ancient Greece, and was known for living a simple life of unflinching moral principle. Legend has it that he once wandered the streets of Athens with a lamp "looking" for an honest man.

Chisholm's example is for mental states that lack a referent, but a similar point applies when there is some referent. For it is also the case that we sometimes seem to see what is not there. Consider an example of representing gone wrong:

> *Raccoon.* In a back alley on a dark night, I see a small four-legged animal approach me, which I believe to be a cat. However, as the creature approaches I realize that it is in fact a raccoon. Prior to my realization, I tokened a concept CAT, which continued to be about cats even though it was caused by a raccoon.

Raccoon is a case of *misrepresentation*: a representation is erroneously tokened by something other than what it represents. Just as the lack of honest men does not undermine Diogenes' ability to look for one, likewise I can readily seem to perceive a cat when only a raccoon is present. Like Diogenes and his tub, I need a cat to pet one, but I don't need a cat to represent one. The usual moral that is drawn from the phenomenon of misrepresentation is that "allowing" for misrepresentation is an important ingredient for intentional content, or the representing relation. Dretske (1988, p.65 emphasis in original) makes the point forcefully:

> it is the power to misrepresent, the capacity to get things wrong, to say things that are not true, that helps *define* the relation of interest. *That* is why it is important to stress a system's capacity for misrepresentation. For only if a system has this capacity does it have, in its power to get things right, something approximating *meaning*.

Accounting for misrepresentation has typically even been taken to be the primary task of theories of intentional content (I return to this topic in Chapter 6). However what is interesting about Chisholm's example is not that Diogenes is somehow in error when he looks for an honest man (though using a lamp was perhaps the wrong method), but rather that the object of Diogenes' attitudes does not need to exist. Likewise, it has been argued that accounting for misrepresentation, or error, is for a theory of representational function rather than a theory of intentional content (Cummins, 1996; Fodor, 1990). I think that error is more a *reflection* of a staple ingredient for intentional content, what Fodor (1990, p.90-91) calls *robustness*: intentional content is common across tokenings of a representation, however the tokenings might be caused. Consider two more cases.

> *Mr.Muscles.* My brother has a cat named Mr.Muscles. He is a handsome tuxedo, with black and white fur, and a somewhat excitable personality. As I sit here thinking of him, I token my concept CAT.

> *Concussion.* Late to get out the door, I rush down the stairs. Losing my footing I trip and fall. At the bottom of the stairs I hit my head, striking my right temple to the floor. The trauma results in me sustaining a mild concussion. Bizarrely, the blow causes repeated tokenings of CAT.

Like *Raccoon*, *Mr.Muscles* and *Concussion* are also examples where my concept CAT is tokened by things other than what it represents—they are what we may call "wild" tokenings (Fodor, 1987). But there are important differences between the cases. When I think of Mr.Muscles no error has occurred, and the tokening of CAT is internally rather than externally caused. When I sustain the concussion, the repeated tokenings of CAT are externally-caused, but they are not obviously *erroneous* for at no point do I make the mistake of thinking a cat is present (even though I cannot get thoughts of them out of my head). In these other cases the content of CAT is robust in Fodor's sense: the content of the concept is the same, whether it is tokened by raccoons, thoughts of a particular cat, or knocks to the head.

So examples of misrepresentation are, I think, relevant to intentional content not because they are a convincing case of error, but rather because they are a convincing illustrations of robustness. The connection to Chisholm's example is that even if the "object" of my representation exists, it need not be present when I token the representation—and it is this fact that is captured by the notion of robustness.[15]

The above examples also suggest different forms of robustness. First, intentional content can be *perceptually* robust: the intentional content of a perceptual representation is common to all tokenings of the representation despite variability in the causes of the sensory inputs that activate the representation. The phenomenon of misrepresentation is a reflection of perceptual robustness. For example, in *Raccoon*

[15] Must wild tokenings of a representation be actual, or can they just be counterfactual? That is, must it be the case that wild tokenings of an internal representation have occurred in the actual world for us to conclude that it is robust? Some insist that there must be actual instances (e.g., Orlandi, 2014), but it is not clear to me why the mere counterfactual possibility is not enough to establish robustness (Fodor, 1990).

my concept CAT exhibits perceptual robustness. Second, intentional content can be *cognitively* robust: the intentional content of a representation is common across tokenings caused by other internal cognitive processes (Fodor, 990b). In *Mr.Muscles* my concept CAT exhibits cognitive robustness. Also, the mental states of Diogenes, as he looked for an honest man through the streets of Athens, plausibly also exhibit cognitive robustness. Lastly, I will say that intentional content is *resilient* to reflect the fact that even when a representation is tokened (directly or indirectly) by extra-psychological means, the content is unchanged (Rey, 1998). In *Concussion* my concept CAT exhibits resilience. Likewise, if the neural realization for CAT was localized and directly stimulated, the content of the concept would remain resilient.

It is plausible that, in general, intentional content is resilient, and depending on the kind of representation, perceptually robust, cognitively robust, or both. Perhaps some internal representations of the visual system cannot be tokened "off-line" in working memory or mental simulation, in which case they are not cognitively robust, since they are only activated by sensory inputs. For example, perhaps the concept CAT that I token when I see a raccoon is not the same as the one I token when I think of Mr.Muscles. For this reason, going forward, I will minimally take perceptual robustness and resilience as constitutive ingredients for perceptual representations in the visual system.

2.4.3 Correctness

In our discussion of robustness we already saw that a notion of error is often associated with mental representation, which points to to another important mental ingredient: that they must have "correctness" conditions. Roughly, these are whatever conditions or states of affairs are picked out by the content of the representations (Burge, 2010, p.38).

I think an argument can be made that given the other ingredients I have specified, a perceptual representation must also have correctness conditions. If the content of an internal representation is perceptually objective, and the representation serves a stand-in function, then it will be correctly tokened whenever its content (what it stands in for) is in fact the case, and has presumably caused the representation to be tokened. When a token of the representation is otherwise caused externally (i.e., wild tokenings), but it is still operating as a stand-in, then the representation will be incorrect, or in error. For example, when I represented the raccoon in the back alley as a cat, I tokened CAT, but given what the concept is about, and the role it plays in my psychology, the tokening would be correct if a cat had in fact been present. Given this argument, if an internal representation is both perceptually objective and robust, it follows that it will have correctness conditions. Since this ingredient is implied by the other ingredients I have listed, I will not emphasize it in what follows.

In summary, I think the following constitutive mental ingredients are necessary for perceptual representation; that is, an internal state is a perceptual representation only if it:

> (i) is a vehicle for original content that is perceptually *objective*, perceptually *robust*, and *resilient*; and (ii) it functions as a perceptually robust "stand-in" with respect to the internal operations of the system.

(i) highlights what I take to be the most important distinguishing ingredients for intentional content relevant to perceptual representation, which, when coupled with (ii), also entails that there are correctness conditions for the representation. Let me say something about why it is important that the state type be a *perceptually robust* stand-in.

As we saw earlier with the example of the compass, the fact that the needle of a compass tracks magnetic North is not enough, by itself, to explain why it is a representational device. For example, a magnetized needle on the leaf tracks magnetic North just as my compass does, but it is not a representation of magnetic North, unless it has the function of indicating magnetic North. So in order to be a representation, the content of the state of a system must somehow be relevant to the functional role of the state within the system. And when it comes to *mental* representation, it must be the case that the *intentional* nature of the content is somehow relevant to the functional role of an internal representation within our psychology (Ramsey, 2007). The most functionally salient of the intentional ingredients is that of perceptual robustness. If the function of an internal representation is to serve as a stand-in—regardless of what causes the representation to be tokened—then its function is to provide a perceptually robust stand-in. Thus (ii) serves to connect the function of a mental representation to its intentional content.

2.5 A Recipe for Perceptual Representation II: Some Special Ingredients

We now have a sense of the mental ingredients necessary for perceptual representation. Are they also sufficient? Operationally, it is tempting to take a perceptual representation to be any internal representation that has all of the ingredients from the last section. But I believe a few more should be included, which are also unique to perceptual representation, and distinguishes it from other forms of mental representation.

In this section I enumerate a few "special" ingredients for perceptual representations, as a species of mental representation peculiar to sensory systems like vision. As is commonly done, I contrast perceptual representations with sensory registers, which have my ingredients for internal representation, but lack the mental ingredients from the previous section. I then discuss some further ingredients for perceptual content, some of which I think are entirely optional.

2.5.1 Internal Representations of Sensory Systems

Something left unspecified by the list of ingredients from the previous section is that a perceptual representation must be an internal representation of a "sensory system".

There is some debate about what the constitutive conditions are for sensory systems, and how these systems are to be identified and individuated. For example, it has been suggested that sensory systems must involve their own form of sensory transduction (Keeley, 2002), or a distinct form of qualitative experience (Matthen, 2015). A separate issues concerns what kind of constitutive conditions should be provided. For example, perhaps "sense modality" should be thought of as cluster-concept (Picciuto and Carruthers, 2013).

These are interesting issues, but I do not think I need to offer a recipe for identifying sensory systems, as a prerequisite for my recipe for perceptual representation. The reason is that vision is clearly a sense modality, if anything is. However, it is important that I distinguish between two kinds of internal representations one can find in a sensory system like vision: those that merely represent proximal properties of sensory inputs, and those that represent the distal world. I mark the distinction, as many others have, as that between a *sensory register* and a perceptual representation (Burge, 2010; Orlandi, 2014). Crucially, while sensory registers are a kind of internal representation of a sensory system, they are non-mental.

First, sensory registers are not perceptually objective in the sense I spelled out earlier. The reason is that if the content of an internal representation is perceptually objective, then it is distal. However, by definition the content of a sensory register is proximal. Second, while sensory registers are plausibly resilient, they are not perceptually robust since (in the case of vision) they only track higher-order properties of the visual stimulus that can be extracted from the retinal image. So long as the light in some part of the retinal produces the appropriate low-level features, a sensory register is activated. For example, internal representations of orientation at a certain spatial frequency, in a particular portion of the visual field, will be a sensory register.[16]

Finally, with respect to function, only perceptual representations provide perceptually robust stand-ins for what they are about. For this reason I think only perceptual representations have correctness conditions. One might try and mark the distinction between sensory registration and perceptual representation by the types of functions they carry out. For example, Burge (2010) argues that sensory registers only serve biological and not representational functions, because he takes correctness conditions to be a prerequisite for having a representational function (for more discussion of this alternate recipe, see Chapter 4). In contrast, I think I can distinguish between sensory registers and perceptual representations, without having to deny that sensory registers can serve a representational stand-in function.

The importance of distinguishing sensory registers from perceptual representations is that in arguing for Conclusion 2, I will need to show why positing perceptual representations, as opposed to mere sensory registers, is required to explain facts about visual object recognition (as we shall see in Chapter 4 and 5).

2.5.2 (Optional) Ingredients for Perceptual Content

So far a perceptual representation is an internal representation of a sensory system that also satisfies my mental ingredients. However it is often claimed that perceptual *content* has other more particular ingredients, which distinguish it from

[16] Though even these internal representations may be perceptually or cognitively robust, if one thinks that cognition penetrates deeply into perception.

the content of other sorts of mental representations. Specifically, it is common to think of perceptual representations as attributing properties to particulars in a manner that is, in some sense, "perceptual" (Burge, 2010; Davies, 1991).

This description of perceptual content invites questions about the nature of these properties and entities, which I do not propose to answer (Nanay, 2015). The reason is that how one answers these questions does not make a difference as to whether an internal representation of a sensory system has my mental ingredients. So long as the entities are in the distal world, and the properties are attributed to them, I am not concerned with whether more specific ingredients are also satisfied. The representations I am interested in are perceptual in so long as they are internal representations of a sensory system, like vision, which have the mental ingredients I presented in the last section. And their content is perceptual so long as it involves attributing a property to a particular. When we come to discuss object vision in Chapter 5, I will have more to say what the relevant properties and particulars might be. But in general I do not think I need to commit myself to anything more specific. In particular, below are two issues I will not take stands on.

First, I am agnostic about the ontological status of the particulars or individuals that vision attributes properties to. Of course they must be *something* that is "out there" in the distal world (Davies, 1991). But I do not think I need to commit myself, for example, to the view that perception attributes properties to ordinary physical objects as some have been inclined to argue (Matthen, 2005; Pylyshyn, 2007). Part of the reason for why I am agnostic is that proponents of this view sometimes take it for granted that we perceive objectively (e.g., Casati, 2015), and

I do not think that the empirical evidence that might be marshaled to support the view supports anything other than the claim that the object of perception is distal.

Second, I am also agnostic about whether the content is nonconceptual; that is, whether the content of perceptual representations is independent of the concepts that a perceiver possesses (Evans, 1982). Similarly I am noncommittal about whether being a perceptual representation (as a kind of state) is dependent on cognitive factors (Heck, 2000). The reason for my agnosticism is that my central case, visual object recognition, is at the border of perception and cognition, and is a central aspect of visual cognition. The representations that underlie the capacity are categorical, and so could be consider either concepts in their own right, or not. If they are not, whether their content or state-type are somehow determined by cognitive factors will very likely depend on what position one adopts regarding the degree to which the visual system is encapsulated from other sense modalities and cognition. But whatever ones standpoint on these issues, they have no bearing on what I wish to argue since whatever position one favors does not seem to have any bearing on whether the representations in question are perceptual (given my recipe). At most these issues would seem to relate to whether we categorize object recognition as occurring at the end of vision or beginning of cognition. But it might also be that object recognition, as a part of both object vision and visual cognition, is simply a penumbra when it comes to arguing about nonconceptual content and state-type individuation.

In summary, I will be relying on the following recipe for perceptual representation in arguing for Conclusions 2 (and hence 3):

> *Perceptual Representation:* any internal representation that: (i) is a vehicle for original content that is perceptually *objective*, perceptually *robust*, and *resilient*; (ii) functions as a perceptually robust "stand-in"; (iii) is internal to a sensory system; and (iv) attributes a property to a distal particular.

(i) and (ii) are just the relevant mental ingredients from earlier, while (iii) and (iv) are the special ingredients I added in this section. As we shall see, showing that all aspects of (i) are satisfied will be the focal point of my arguments for Conclusion 2, as developed in Chapters 4 and 5.

2.6 Missing Ingredients? Some Anti-Representational Challenges

Given the way I have laid out my recipes for internal and perception representation, my arguments for Conclusion 1 and 2 can fail for two reasons:

> (1) The representations posited in typical explanations of some visual phenomenon of interest are missing ingredients for internal representation.

> (2) The representations posited in typical explanations of some visual phenomenon of interest are missing ingredients for perceptual representation.

If true of the information-processing framework in general, (1) Undermines my first two conclusions, while if true of explanations of object recognition, (2) only undermines Conclusion 2. In this section I review three anti-representationalist arguments. All of these arguments were developed to target the utility of mental representations as a posit in cognitive science. However, in light of my recipes for internal and perceptual representation, the first two arguments can be thought us reasons for endorsing (1), while the third provides a reason to endorse (2). Thus I treat them as general sorts of "challenges" that I will need to respond to.

2.6.1 The Informal Challenge

One way to undermine both arguments is to show that even some of my generic ingredients are not satisfied by internal states of the visual system, as claimed by Chomsky (1995, p.52):

> There is no meaningful question about the "content" of the internal representations of a person seeing a cube under the conditions of the experiments . . . or about the content of a frog's "representation of" a fly or of a moving dot in the standard experimental studies of frog vision. No notion like "content" or "representation of" figures within the theory.

Chomsky denies that there is *any* notion of representational content that is of explanatory importance to vision science, and appears to reject the very idea that representing is a kind of relation. In effect all one needs to posit in order to explain how we see are ersatz representations. One motivation for Chomsky's view is the claim that any intentional idioms that are used in vision science are *informal*. A vision scientist might:

> speak of "misperception" in the case of the person or frog in the experiments, though perhaps not when a photoreceptor on a street light is activated by a searchlight rather than the sun ... But these usages are *on a par* with an astronomer warning that a comet is aiming directly toward the Earth, implying no animist, intentional physics.(Chomsky, 1995, p.53 my emphasis)

Intentional descriptions of representations are informal, and like informal intentional descriptions of phenomena in other branches of science, such descriptions are inessential, and hence dispensable, to explanations that posit some form of representation.

Although directed at appeals to mental representation, the heart of the arguments seems to be the denial that any notion of representational content is required for explanations in cognitive science to succeed. So if Chomsky's argument succeeds, it undermines all of my conclusions in a fundamental way. The argument presents what we may call *the Informality Challenge*: To meet it I must show that, minimally, a content component is indeed required as part of the information-processing framework in vision science. And this requires a convincing argument for why reference to representational content does not reflect a mere intentional gloss, as one sometimes finds in other areas of science.

2.6.2 The Instrumental Challenge

A different style of argument, which also presents a challenge to my arguments, rests on the claim that the notion of representational content employed in typical explanations of a psychological phenomena satisfy *other* ingredients, which run counter to those I have offered. Egan (1992, 1995, 1999, 2010) has proposed a particular *explanatory* function for representational content in computational models in cognitive science, according to which the content is derived. Egan (1995, p.84-85) agrees with Chomsky that intentional descriptions are often used informally to describe visual processes, but such processes are also described in *formal* terms, which requires an intentional description. On the one hand we have some computational model, which can be described in purely formal terms; and on the other, there is the phenomenon of interest. According to Egan, providing a distal interpretation of the model links it to the phenomenon, which makes the model explanatory.

> The questions that antecedently define a psychological theory's domain are usually couched in intentional terms. For example, we want a theory of vision to tell us, among other things, how the visual system can detect three-dimensional distal structure from information contained in two-dimensional images. A *characterization* of the postulated computational processes in terms of distal objects and properties enables the theory to answer these questions . . . It is only under an *interpretation* of some of the states of the system as representations of depth and surface

> orientation that the processes given a mathematical characterization by a computational theory are revealed as *vision*. (Egan, 2010, p.256 my emphasis; cf. Egan, 1995, p.189)

Egan acknowledges that intentional descriptions might point us toward a phenomenon of interest, but she distinguishes between intentional content and what I will call *instrumental content*.[17] It is the latter that plays an important explanatory role in explanations of visual phenomena, since it serves to link computational models to the phenomenon of interest, so that the models become an explanation of the phenomenon (Egan, 1992, p.452); that is, attributing instrumental content provides a "pragmatically motivated gloss" of a computational model (Egan, 2010, p.259).

The ingredients for instrumental content are quite different from those I specified for internal and perceptual representation. First, serving an explanatory role for a theorist is an entirely different notion from a representational function for the system of interest, and so it is no surprise that instrumental content is clearly derived: it is only under an interpretation of a computational model of a system that it has instrumental content, and interpretations are interest relative and context sensitive with respect to our explanatory goals (Egan, 1999, p.186). Second, instrumental content can be distal, and directed at property instantiations, but it is a researcher's explanatory interests that warrant the distal, rather than a proxi-

[17] My recipe for perceptual representation is close to the position she calls the "Essential Distal Content View" (Egan, 2010), and "hyper representationalism" (Egan, 2012). However, Egan lumps together my recipe with the requirement that intentional contents are essential to, and serve to individuate, representational states. But I exclude the latter requirement as part of my recipe.

mal, interpretation (Egan, 1995, p.197). As Egan emphasizes, endorsing the idea of instrumental content amounts to the rejection of the sort of ingredients that I have in mind, which includes content being original (e.g. Egan, 2010, p.259). According to Egan, ascribing instrumental content in the explanations of cognitive science has the same function as interpretations utilized in other domains of natural science that use computational models (Egan, 1992, p.444).

Egan's position conflicts with my recipes for internal and perceptual representation because she maintains that the only notion of representational content in cognitive science, which does actual explanatory work, has derived, instrumental content. Hence if her argument succeeds, she undermines Conclusions 1 - 2 (and therefore 3). Thus her argument presents what I call the *Instrumental Challenge*. In order to overcome it, I need to show that the sort of representational content that is appealed to in the information-processing framework is not simply the instrumental content that is attached to computational models in order to make them explanatory.

2.6.3 The Job Description Challenge

The two previous challenges in effect deny that researchers really have notions of internal or perceptual representation in mind when it comes to the "representations" that they posit in their explanations. As we have seen, this is largely because they reject the idea that intentional descriptions in the explanations of cognitive science imply the positing of mental representations. A complementary form of argument grants that researchers might have something like a notion of mental representation that they mean to employ in their explanations, but under inspection, their posits fail to have the right ingredients. Specifically, purported mental representations do not have any representational function tied to properties of intentional content. And without such a function, the posits lack a necessary ingredient of mental representations. Ramsey (2007, p.27) refers to the need for a notion of representational function (which is tied to intentional content) as the "Job Description Challenge":

> There needs to be some unique role or set of causal relations that warrants our saying some structure or state serves a representational function. These roles and relations should enable us to distinguish the representational from the non-representational and should provide us with conditions that delineate the sort of job representations perform, *qua* representations, in the physical system. I'll refer to the task of specifying such a role as the "*job description challenge*." What we want is a job description that tells us what it is for something to function as a representation in a physical system.

Absent a job description, that is, a description of the representational function of an internal state that (purportedly) has intentional content, the posited internal state is not a mental representation. Ramsey argues at length that while there are some notions of representation in cognitive science that seem to meet his challenge, the notions that are popular in contemporary research do not.[18]

For example, one notion of representation sometimes found in vision science is what Ramsey calls the "receptor" notion. This notion, which is very similar to what I am calling sensory registers, fails the Job Description Challenge. The receptor notion has its origin in early single-cell recording work on the retina of various model organisms (e.g., Barlow, 1953; Lettvin et al., 1959), and similar views have been defended by philosophers (e.g., Dretske, 1988). According to Ramsey (2007, p.120) the core of the receptor notion is that: "because a given neural or computational structure is regularly and reliably activated by some distal condition, it should be regarded as having the role of representing (indicating, signaling, etc.) that condition." However the receptor notion is not obviously a notion of *mental* representation, since it only requires that an internal state of my visual system is reliably activated by a state of the world:

> There are several *non-representational* internal states that must, in their proper functioning, reliably respond to various states of the world. Our immune system, to take one example, functions in part by consistently reacting to infections and insults to our bodies, yet no one suggests that any given immune system response (such as the production of antibodies)

[18] For replies to Ramsey's arguments see Morgan (2014) and Shagrir (2012). For my part, I think Ramsey offers an incomplete recipe for mental representation, and so even the cases that he thinks meet his challenge might indeed fail. The reason, briefly, is that the only distinguishing ingredient he offers for intentional content is its originality. But as we have seen above, this is better seen as an ingredient for internal representation, and other ingredients need to be provided.

> has the functional role of representing these infections. (Ramsey, 2007, p.125)

Having the function of reliably being caused by a state of the environment is not sufficient to make an internal state of a system a mental representation, even if it is a neural state. Hence, the receptor notion, as described, fails the job description challenge. Ramsey's style of argument could be equally successful at undermining my argument for Conclusion 2. Meeting the job description challenge means showing that the internal representations posited by some explanations of visual phenomena have my ingredients for perceptual representation, and that includes a distinctive representational function—in my case, they must function as a perceptually robust stand-in.

2.7 Conclusion

In this chapter I outlined my argumentative strategy for my main conclusions about the roles of different notions of representation in explanations in vision science. Carrying out the strategy required stocking up with a set of recipes for different notions of representation. The cupboard is no longer bare. We now have at our disposal notions of internal and perceptual representation. It is now time to put these recipes to use.

Chapter 3: The Contents of the Information-Processing Framework in Vision Science

3.1 Introduction

In this chapter I defend Conclusion 1: a notion of internal representation is indispensable to explanations within the information-processing framework in vision science. The notion of "internal representation" I introduced in the last chapter has two key ingredients: that the content of the representation is original, and that the representation has a "stand-in" function. Thus, to defend my first conclusion, I need to show that representations in the information-processing framework have these crucial ingredients, and compare my argument to the three anti-representationalist challenges from the last chapter.

The most well-known characterization of the information-processing framework is laid out in Marr's classic *Vision* (Marr, 1982). What I propose to argue in this chapter is that a notion of internal representation is indispensable to Marr's characterization of the framework, and in so far as his approach can be taken as representative of the field at large, this provides an argument for my first main conclusion. One might worry that focusing so narrowly on the work of one individual substantially weakens my case. However there are three reasons why this focus is warranted.

First, some of the anti-representationalist challenges from the previous chapter have largely relied on Marr's work as a case study (Chomsky, 1995, 2000; Egan, 1992, 1995, 1999, 2010). Thus, I can meet their challenges by showing that they fail with respect to Marr's work. Second, contemporary research on visual object recognition was born from, and remains largely indebted to, Marr's work on the subject (Marr and Nishihara, 1978). Third and finally, Marr's work continues to have a substantial influences on vision science, and cognitive science more generally. This is evinced by the recent special issues of the journals *Perception* (2012) and *Topics in Cognitive Science* (2015) discussing his legacy. So for these three reasons, I think it is reasonable to treat Marr's approach, in the present context, as representative of the commitments of the information-processing framework more generally.

Since it was first published, interpreting Marr (1982) has been something of a cottage industry in philosophy, especially when it comes to the role that representational content (of some sort) plays in his "philosophical" approach. Marr was careful to emphasize two things in developing his approach: (i) the fundamental duality in vision between "representation" and "process"; and (ii) that investigating this duality required multiple levels of analysis. My own interpretative conclusion is that with respect to (i) Marr clearly had in mind a notion with all my ingredients for internal representation, which alo can be characterized at multiple levels, in accordance with (ii). The notion is constitutive of, and hence, indispensable to, his approach. Attributing a content-component to Marr's approach is surprisingly controversial. But as we shall see, I think it is in fact hard to resist.

The rest of the chapter is structured as follows. In Subsection 2, I make some preliminary remarks about the information-processing framework in general, and alternative frameworks in vision science. In Subsection 3, I review the core of Marr's information-processing framework: how he characterized the notions of "representation" and "process", and his well-known three levels of analysis. In Subsection 4, in order to illustrate Marr's approach in action, I review his work on edge detection (Marr and Hildreth, 1980). In Subsection 5, I argue that Marr's approach requires a "content-component", which I position (as others have) at the highest of his three levels. Marr's notion of representation, I argue, has the ingredients for internal representation, and is indispensable to his approach. In Subsection 6, I evaluate the three anti-representationalist arguments with respect to Marr's approach. Subsection 7 concludes the chapter.

3.2 The Structure of the Framework and the Story of Vision

Before delving into the details of Marr's approach, I want to address some general issues relating to the information-processing framework in vision science. In this section I offer some minimal characterizations of key notions employed in the framework. I then briefly discuss alternative frameworks in vision science that are, in one way or another, anti-representationalist.

3.2.1 Clarifying some Aspects of the Information-Processing Narrative

There are three aspects of the "information-processing" framework that are rarely well-articulated by vision scientists: the notions of "information" and "computation", and the explanatory-style of the framework. This lack of clarity also applies to Marr's work. So before proceeding, I would like to stipulate characterizations of each of them that I think are fairly minimalist in their assumptions. These will be the characterizations I will be assuming in the discussion to follow.

Information. When talking of "information" it is not uncommon for vision scientists to reference communication theory (Shannon, 1948), and the formal notion of information as the reduction of uncertainty (or entropy, depending on one's interpretation) about the state of some system. It is also common for them to talk of the related notion of mutual information, which is a measure of the mutual dependence between two random variables. Methodologically referencing these measures makes sense, since both are quite useful when it comes to statistical analysis and computational modeling.[1] However it is doubtful that this is the notion that they have in mind when talking of "information-processing" by the visual system. Rather, a more plausible assumption is that they have in mind the notion of *natural* information.[2] This is the sort of information that one event carries about another

[1] As just one example, mutual information is the standard measure that is used for motion correction during fMRI preprocessing.

[2] The inspiration comes from Grice's (1957) distinction between natural and non-natural meaning (Piccinini and Scarantino, 2010).

event that it reliably co-varies with. For example, it is this kind of information that tree rings carry about the age of a tree, smoke carries about fire, the whistle of a kettle carries about the temperature of the water it contains, the mercury in a thermometer carries about ambient temperature, or the needle of a compass carries about a magnetic pole. In each of these cases, due to the dependence of one event on another event, the former is informative about the occurrence of the latter.

While there are a variety of ways to understand the notion of natural information, the most familiar is simply that of causal covariation.[3] Thus, going forward, I will assume that all talk of "information-processing" is of natural information. In other words, it is this sort of information that is carried by the retina, and that the visual system "processes".[4]

Computation. There are myriad notions of computation that can (and have) been utilized in vision science, and cognitive science more generally.[5] Others have worked through the nuanced relationship between these different notions (see e.g., Maley, 2011; Piccinini and Scarantino, 2010). For present purposes, I want to assume a broad notion that brings with it few substantive commitments. For this purpose,

[3] For example, it has been proposed that the relationship is probabilistic (Dretske, 1981; Scarantino, 2015; Skyrms, 2010), nomic or law-like (Dretske, 1988; Fodor, 1990), or counterfactual (Cohen and Meskin, 2006).

[4] At present this might seem like an innocuous point, but it will have a good deal of significance when we come to discuss the Transformational Thesis again in Chapter 6.

[5] For example, there are various dichotomies that can be drawn such as abstract vs concrete, digital vs. analog, discrete vs. continuous, and specific types that can be identified such as algorithmic, symbolic, and neural.

a notion of "generic" computation, which I owe to Piccinini and Scarantino (2010, p.239), will do nicely:

> We use 'generic computation' to designate *any process whose function is to manipulate medium-independent vehicles according to a rule defined over the vehicles*, where a medium-independent vehicle is such that all that matters for its processing are the differences between the values of different portions of the vehicle along a relevant dimension (as opposed to more specific physical properties, such as the vehicle's material composition).

I think this notion has sufficient generality to capture a good deal of what vision scientists have in mind when they talk of computation. Crucially, as with the notion of internal representation from the previous chapter, it involves no commitment regarding what the computational architecture of the visual system is like, and whether it is classic or connectionist/distributed. Defined in this broad sense, the sort of information-processing exhibited by the visual system (i.e., the processing of natural information) is minimally computational in that it qualifies as a form of generic computation (Piccinini and Scarantino, 2010, p.243).

Explanation. The last aspect of the framework requiring clarification is the issue of what *sort* of explanations it offers. The two most obvious alternatives are that the explanations of the framework are functional (Cummins, 1983; Fodor, 1968) or mechanistic (Bechtel, 2008).[6] A good deal has been written about how these two

[6] How to characterize mechanisms, or mechanistic explanation, is itself a source of debate.

forms of explanation do or do not feature in psychology and neuroscience. When it comes to characterizing research in these fields, and in cognitive science, there are in fact two issues first, are these two types of explanation distinct; and if so, which one provides a superior characterization of the target explanations (e.g., in visual neuroscience); And second, if they are not distinct, what is the relationship between them? For example, one view is that functional analyses of different sorts are simply incomplete forms of mechanistic explanation, or "sketches" (Piccinini and Craver, 2011).[7]

These are interesting issues in the philosophy of science, and the information-processing framework in general, and Marr's work in particular, provide fertile ground for investigating them. However, I do not think I need to take a stand on either issue. Addressing them requires determining how we should relate explanations at multiple different levels in cognitive science, and the extent to which these levels are distinct or autonomous from each other. This is also true when it comes to interpreting Marr's work. But all that is important for my purposes is that there are these levels, and that at least within Marr's work, a content-component is indispensable to one of them. For example, I have no interest in evaluating whether or not Marr's levels of analysis are distinct, such that each offers a different perspective

One influential characterization is as follows: "Mechanisms are entities and activities organized such that they are productive of regular changes from start or set up to finish or termination conditions (Machamer et al., 2000, p.3). In which case, mechanistic *explanations* offer descriptions of mechanisms, or "schemas", consisting of various entities, activities, and their organization that account for how the phenomenon of interest is produced.

[7] Another issue is whether some computation is *itself* always mechanistic (Piccinini, 2007).

to understanding information-processing mechanisms (Bechtel and Shagrir, 2015), or that some of the levels should be lumped together (Piccinini and Craver, 2011).

So in summary, I assume that minimally the framework (and Marr's approach to it) posits the processing of natural information, which is computational in a generic sense, and that the explanations that are offered by the framework are functional and/or mechanistic.

3.2.2 Tall Tales? Alternative Tellings of the Story

Although the information-processing framework is the dominant research program in vision science, there are alternative tellings of the story of vision in which representations do not appear. One might reasonable ask why have I have chosen not to include them as challenges to my arguments. Let me say a little about these alternative tales, and why I am not asking you to read about them.

Most famously Gibson (1950, 1979) held that visual perception requires no internal representations because their are higher-order properties of retinal stimulation that uniquely map onto properties in the distal world. In which case, there is no need to posit internal representations and processes that try to extract and build upon information latent in the retina. Instead the goal of vision science should be to discover and characterize the higher-order properties of the visual input that allow us direct access to the distal world. Although today this is a minority view among vision scientists, there are still a devoted few who adopt some version of Gibson's direct realist/ecological psychology framework.[8] Some of these views follow fairly

[8] A good number of them are affiliated with the Center for the Ecological Study of Perception

directly in the footsteps of Gibson (e.g., Cutting, 1982; Turvey et al., 1981), while others consider his ideas more of an inspiration. Two examples are sensorimotor approaches, which characterize vision as an active exploration of the world that is mediated by sensorimotor contingencies (O'Regan and Noë, 2001), and "embedded" views, which hold that the visual system is organized in a manner that reflects regularities in the environment without representing them (Orlandi, 2014). Both approaches share commonalities with Gibson's approach, and reject the need for positing "representations" in any interesting sense.

While these approaches present important challenges to the information-processing framework in vision science, I do not intend to evaluate their relative virtues. The reason for this is that the focus of my project is on whether certain notions of representation are indispensable to the explanations *within* the information-processing framework that purport to posit them—not to determine which framework offers the best explanation of various visual phenomena. Thus the challenges from the previous chapter can all be seen as "internal" to the information-processing framework, in contrast to the "external" challenges of alternative frameworks in vision science. I consider these external challenges to be outside the scope of the present project.[9] Still, even if one is interested in adjudicating between research programs in

and Action, at the University of Connecticut.

[9] A seemingly more relevant challenge is provided by those within the information-processing framework that have argued against the Distal Object Thesis, and perceptual objectivity (Mark et al., 2010). However these views are largely motivated by considerations of whether perceptual objectivity is fitness enhancing, and are not obviously targeted at a plausible form of objectivity (Cohen, 2015). For this reason, I consider the issues they raise to be somewhat orthogonal to my

vision science, addressing internal challenges has values. The reason is that getting clear on internal issues about the role that different notions of representation must play within the information-processing framework puts us in a better position for evaluating its explanatory merits relative to competing frameworks in the field.

Having dispensed with these preliminaries, I now turn to the exposition of Marr's influential ideas.

3.3 Origins of the Framework: Marr's Vision of Vision

The general form of the information-processing framework in vision science had its first clear formulation in Marr (1982). For this reason it remains an important touchstone for those interested in characterizing the representational commitments of the framework. Marr (1982, p.3) begins with the question: "What does it mean, to see?" The intuitive answer, "to know what is where by looking", provides the point of departure for Marr's discussion:

> In other words, vision is the *process* of discovering from images what is present in the world, and where it is.
>
> Vision is therefore, first and foremost, an information-processing task, but we cannot think of it just as a process. For if we are capable of knowing what is where in the world, our brains must somehow be capable of *representing* this information—in all its profusion of color and form, beauty, motion, and detail. The study of vision must therefore include

concerns.

> not only the study of how to extract from images the various aspects of the world that are useful to us, but also an inquiry into the nature of the internal representations by which we capture this information and thus make it available as a basis for decisions about our thoughts and actions. This duality—the representation and the processing of information—lies at the heart of most information-processing tasks.(ibid)

According to Marr, the fact that vision is an information-processing task implies it is both a process, and representational.[10] Marr's great theoretical insight was the recognition that in order to understand how a system performs an information-processing task, we must also understand the system at many levels: in terms of *what* it is doing and *why*, but also *how*, and with what kind of architecture

> For the subject of vision, there *is* no single equation or view that explains everything. Each problem has to be addressed from several points of view—as a problem in representing information, as a computation capable of deriving that representation, and as a problem in the architecture of a computer capable of carrying out both things quickly and reliably. (Marr, 1982, p.5)

In general, Marr held that understanding any complex system requires multiple levels of analysis. For example, a developing embryo can also be analyzed at multiple levels. But when the system is performing an information-processing task, one

[10] Marr never defines the notion of information he has in mind. As stated above, I am assuming a notion of natural information.

must *also* appeal to the notions of representation and process (Marr, 1982, p.20). Thus Marr's "philosophical" approach can be seen as having two core elements: (i) the duality between representation and process; and (ii) levels of explanation, or analysis. In this section I provide expositions of both (i) and (ii). Because of the controversy in interpreting Marr, I make ample use of quotations in support of my own characterization of his ideas.

3.3.1 The Central Duality: Representation and Process

Marr (1982, p.20-21 emphasis in original), emphasis in original) defines a representation as follows:

> A *representation* is a formal system for making explicit certain entities or types of information, together with a specification of how the system does this. And I shall call the result of using a representation to describe a given entity a *description* of the entity in that representation (Marr and Nishihara, 1978).

Marr illustrates his notion of a representation using numerals. Arabic, binary, and Roman numerals are all different representations for describing numbers. Different numeric systems make more or less explicit different information about a number. For example, Arabic numbers make it easier to determine whether a number is a power of 10, while binary makes it easier to determine whether a number is a power of 2 (because the former is in base 10, and the latter base 2).

A couple of ideas are invoked in Marr's definition. First, by "formal scheme" Marr means a set of symbols with rules for combining them: "To say that something is a formal scheme means only that it is a set of symbols with rules for putting them together—no more no less" (Marr, 1982, p.21). So for Marr 'representation' refers to a formal scheme for constructing multiple instances of what I would call individually a "representation". To disambiguate, I will subsequently use "representational scheme" to refer to the sort of formal system of encoding (e.g., the arabic numerals), which Marr calls a "representation". Also required then are "symbols" with rules for combining them. This use of 'symbol' also invites some ambiguity, since it is also commonly used to refer to internal representations over which computational operations are defined. Therefore, following some of Marr's later terminology, I will instead refer to these symbols as the *primitives* of a representational scheme (see e.g., Marr, 1982, p.37 Table 1-1).

Second, Marr's definition invokes the idea that a representational scheme serves to "make explicit" information extracted from the retina, and a story is needed of how this is done. Marr's notion of making information explicit comes from Marr and Nishihara (1978), who appeal to Marr's earlier "principle of explicit naming", which is supposed to apply to symbolic (i.e., computational) processes (Marr, 1976, p.485). The idea behind the principle is that when a data set is to be manipulated it should first be given a name: "This forms the data into an entity in its own right,. . . and allows other structures and processes to refer to it" (ibid). With his principle Marr is clearly invoking the common notion in cognitive science of a representation functioning as a "stand in", which was introduced in the previous chapter as an ingredient for internal representation.

Third, a "description" in Marr's sense appears to be equivalent to what is more commonly now called an "encoding" of information. Thus an internal representation functions as a stand in for some entity that it encodes information about, at least for further systems that can "read out" the description, or "decode" the information. For example, for someone who does not understand the Roman numeral system then the string 'LVII' fails to stand in for the number fifty-seven—for them the information about number conveyed by the string is implicit (Kirsh, 1992). Thus by providing a set of primitives, and rules for combining them, a representational scheme allows a system to construct descriptions of things, and in so doing, they make explicit information that would otherwise be left implicit.

As Marr points out, his notion of a representational scheme is quite general. It is the idea that we can: "capture some aspect of reality by making a description of it" (Marr, 1982, p.21). Marr's view was that stages of visual processing amount to mappings from different representational schemes, each of which make different information about the retinal image, or visible world, explicit. However, crucially unlike the cash register, there is no reason to think Marr believes that the contents of a representational scheme are derived (a point I return to later).

As with the notion of representational scheme, Marr (1982, p.22) also acknowledges that the notion of a process is quite broad: performing addition is a process, but so is making a cup of tea. Marr's focus was the notion of a process in connection to a system performing an information-processing task, which Marr claimed required an analysis at many different levels. At the most abstract level there is the *computational theory* of a process, which has two components: (i) a specification of *what* operation is being performed by the device, and (ii) an account of *why* it is that the device is performing the operation.[11] Call these the "what"- and "why"-components of the computational theory. For Marr, the why-component includes specifying various constraints that uniquely specify the operation carried out by the device in order to fulfill its information-processing task.[12]

Marr uses the example of a cash register to illustrate the idea of a computational theory. Cash registers perform arithmetic, and addition and subtraction. This is the what-component of the computational theory for the cash register. But the reason it performs arithmetic, as opposed to another mathematical operation, is that it is designed to calculate the appropriate amount of money that needs to be exchanged for a purchase. Certain rules seem fair and appropriate for combining prices and determining the money owed by a consumer. For example: if I buy nothing, I owe nothing; the order of goods does not impact the total sum; paying for different items individually does not impact the total sum; and if I return an item, I

[11] Like with the notion of information referenced by "information-processing", Marr never defines the notion of computation he has in mind. The "what"-component of the computational theory is seemingly just whatever operations or mathematical function an information-processing system carries out. So it seems he has in mind the intuitive idea that when a system carries out a mathematical operation, or logical operation, it "computes". In general, I believe this comports with my assumption of the notion of generic computation earlier.

[12] Some have claimed that the computational theory only specifies what is computed, but not why. I agree with Shagrir (2010) that Marr's actual writing leaves little doubt that the computational theory includes both a what- and why-component.

get a refund, paying nothing. It is a theorem that these constraints define addition, and hence it is appropriate that a cash register computes addition. These rules for transactions involving the exchanging of goods of predetermined monetary value constitute constraints on the operations that should be carried out by a cash register. These constraints in part constitute the why-component for the computational theory of a cash register.

For Marr, the computational *theory* of a cash register is not a device, but really a blueprint of what we want such a device to do, and why. Marr's second level has two components: a representational scheme for the inputs and outputs of the device, and an algorithm for achieving the transformation from the input to output in line with the what-component of the computational theory. For the cash register, the input and output utilized the same representational scheme (strings of numerals from the Arabic numeric system), but the two need not be the same. In general, this second level specifies *how* the device carries out the operation specified by the computational theory. The specifics can vary, as different representational schemes can be used, and for a specific representational scheme, different algorithms might be possible. Lastly, there is the hardware implementation, the physical substrate in which the device is realized. While the same representational scheme and algorithm can be implemented in different physical systems, it is also true that some algorithms are better suited for certain physical systems.

3.3.2 The Three Levels of Explanation

In summary, characterizing a device that performs an information-processing task involves answering questions at three levels (after Marr, 1982, p.25 Fig.1-4):

> **Level 1**: *Computational Theory*. *What* is the goal of the computation performed by the device, *why* is it appropriate to the task, and what is the logic of the strategy by which it can be achieved?
>
> **Level 2**: *Representation and Algorithm*. How can the computational theory be implemented? What is the representational scheme for the inputs and outputs, and what is the algorithm for the transformation from an input representation to an output representation?[13]
>
> **Level 3**: *Hardware Implementation*. How can the representational scheme and algorithm be physically realized?

It is important to see that Marr's well-known three levels are a product of how we understand information-processing tasks in particular, which constitutively involves a representational scheme and process. While any device performing an information-processing task can be understood at all three of these levels, particular explanations might be directed at a particular level based on the phenomenon of interest.

[13] Marr sometimes refers simply to the "algorithmic" level, and following this some authors fail to associate a representational scheme with this level of explanation. But such an omission distorts the structure of Marr's framework.

In articulating and applying his multi-level approach, Marr focused on what he took to be the unappreciated importance of the computational theory. His criticism of previous research was that it made little distinction between questions of *what* a device is doing and *how* it is doing it. Marr makes the elegant comparison that trying to understand vision by simply studying the behavior of individual neurons is like trying to understand bird flight by only looking at feathers—we must also have an understanding of the principles of aerodynamics, which put constraints on (for example) what sorts of limbs could achieve lift, in order to understand why wings and feathers can perform the task of flight.

3.4 Seeing Marr's Vision in Action

In applying his approach Marr characterized vision as a process that starts with retinal inputs, and extracts a description of the external world that is useful to the observer. The nature of the input representational scheme is already given: it is the retinal image itself, or rather the matrix of light intensity values as detected by photoreceptors in the eyes. According to Marr, it makes sense to characterize the retinal image as a representational scheme: what it represents (i.e., the information it makes explicit) are the intensity values at each point in a retinotopic array, which he denoted as $I(x, y)$, for coordinate (x, y). Marr also made the further simplifying assumption that we ignore color, and treat I as in effect a gray scale pixel array (Marr, 1976).[14]

[14] Marr uses the label 'image' to refer to I, though he clearly has in mind an abstract characterization of the intensity levels detected by the photoreceptor of the retina. I use the term "retinal

While the input representational scheme is obvious, the output representational scheme—what vision is *for*—is less obvious. Marr (1982, p.36) himself was convinced that the "chief" information-processing task of vision was to extract information about the 3-D shape of objects in the distal environment for the purpose of recognition. All other intermediary stages in the mapping from the input representational scheme, the retinal image, to the output representational scheme, one for shape, were seen as important "service" tasks.[15] Marr separated visual processing into four stages.

1. The retinal image. A light intensity array, reflecting stimulation of the photoreceptors.

2. The Primal Sketch. Represents information about the retinal image, specifically changes in intensity values and their retinotopic position and organization.

3. The 2 1/2 Sketch. A representation scheme for the orientation and surface depth of visible surfaces in a viewer-centered coordinate frame.

4. The 3-D Model Representation. An object-centered representational scheme for the 3-D structure and organization of visible objects.

In anticipation of later discussion, notice that for Marr, each of these representational schemes are partially characterized by what they represent: light intensity,

image", since he also uses 'image' to refer to natural images, resulting in some ambiguous passages.

[15] the distinction between "chief" and "service" tasks is from Shapiro (1997).

intensity changes, visible surfaces, and the 3-D shape of visible objects. Figure 3.1 depicts the four stages of visual processing, and the purpose of each representational scheme (after Marr, 1982, p.37 Table 1-1).

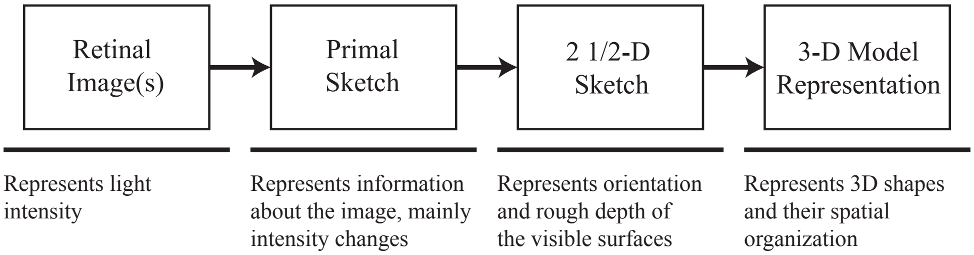

Retinal Image(s)
Primal Sketch
2 1/2-D Sketch
3-D Model Representation
Represents light intensity
Represents information about the image, mainly intensity changes
Represents orientation and rough depth of the visible surfaces
Represents 3D shapes and their spatial organization

Figure 3.1: Stages of Vision for Deriving Shape Information.

Marr developed highly detailed proposals for each of these stages of visual processing. As an illustration of how Marr applied his framework at a particular stage, I will review his work on edge detection in the primal sketch (Marr and Hildreth, 1980), since it has been widely discussed by philosophers. It has also been relied on extensively by Egan (1992, 1995, 1999, 2010) in developing the Instrumental Challenge.

3.4.1 Edge Detection

Marr (1976) separated the construction of the primal sketch into two parts.[16] First, a description is built of the intensity changes found in the retinal image, using

[16] The notion of "sketch" for Marr is a result of his conviction that the visual system basically functions to construct the equivalent of elaborate line drawings. For example, in Marr (1976), where he introduces the notion of the primal sketch, he refers to different "drawings" made by the visual system.

a representational scheme of primitives such as “edge-segments”, “bars”, “blobs”, and “terminations”. Tokens of these primitives occur at different retinotopic scales, locations and orientations. Marr (1976) called this first representational scheme the *raw* primal sketch. Second, larger scale geometric relations between the intensity change primitives are grouped together forming the *full* primal sketch. For example, in Figure 3.2 there are local bars, which we also perceive as grouping to form a rough circle surrounded by a ring (after Marr, 1982, p.92 Figure 2-33a). Under Marr’s account of early vision, the tokening of bar primitives at different locations and orientations is achieved by the raw primal sketch, while the grouping of these bar primitives into a circle and ring is achieved by the full primal sketch.

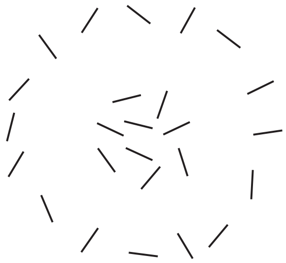


Figure 3.2: Grouping of Primitives.

Marr and Hildreth’s (1980) theory of edge detection was intended to provide an account of the construction of the raw primal sketch. Among philosophers, their theory is perhaps the most widely discussed example from Marr’s work, providing a central case to illustrating elements of his approach in general, and the Level 1 computational theory in particular. Despite the frequent reference to the theory, I know of no case where a discussion of the theory by a philosopher includes a detailed explication of how the theory is intended to work. Getting clear on the mechanics of the theory will help to evaluate both my argument to follow and the anti-representationalist challenges, some of which have been motivated using Marr and Hildreth's theory.

The construction of the raw primal sketch itself has two parts: (i) the *detection* of intensity changes, and (ii) the *representation* and description of these intensity changes via the tokening of different representational primitives (blobs, edges, etc.). Marr and Hildreth's account of how the retina detects intensity changes was motivated by two observations, the first of which relates to the demands of image processing in computer science.

First, an issue when it comes to filtering natural images is that intensity changes occur at widely different spatial frequency scales.[17] For example, when filtering a natural image (to remove noise or distortions) one must set the filter to different scales. Marr and Hildreth suggested that the retina faces the same challenge when detecting intensity changes in the retinal image I, which can also be characterized, abstractly, as a pixel matrix. For the retina must "smooth" the intensity values of the retinal image in order to detect changes at different spatial frequency scales. Furthermore, the right filter should be spatially localized (much like in image processing), since the objects in the world that give rise to intensity

[17] Spatial frequency is a measure of the units of periodicity for some measure of spatial distance. In vision science, the measure of distance is usually defined either in terms of the pixels of an image, or in terms of degrees of visual angle (which is determined by viewing distance of an observer).

changes (such as changes in illumination, orientation, or surface reflectance properties), are spatially localized, and thus the intensity changes they produce will also be retinotopically local. Thus filtering of the retinal image should arise from a smooth average of nearby points in I.

The second important observation is that if we think of an intensity change in an image as a sigmoid function (an "S" curve), then this gives rise to a *zero-crossing* in the second derivative of the function for intensity across I at the very point where there is a change in intensity (Figure 3.3). Thus a filter of a natural image showing only the zero-crossings will do a good job of extracting the intensity changes in the image; in fact, Marr and Hildreth (1980, p.192) proposed to *define* an intensity change as a zero-crossing, so that the task of detecting an intensity change reduces to that of detecting a zero-crossing of the second-derivative with the appropriate direction.

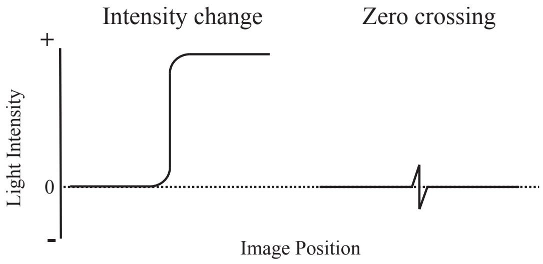


Figure 3.3: Intensity Change and Zero Crossing.

Together, these two observations were intended to provide *constraints* on any adequate filter for detecting intensity changes in I: first, the operator should operate at different spatial frequency scales; and second, it should be a differential operator. As constraints these observations form part of the why-component for Marr and

Hildreth's computational theory of edge detection. Marr and Hildreth argued that the best operator that satisfies these requirements is the filter $\nabla^2 G$, or the "Laplacian of the Gaussian". The Operator has two components, which I will describe in turn. ∇^2 is the 2-D Laplacian operator, which is the sum of the second partial derivatives for each argument of a two-place function $f(x, y)$:

$$\nabla^2 f(x,y) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \tag{3.1}$$

Where ∂ is the symbol for a partial derivative (the two fractions in Equation 1 are the second partial derivatives for the arguments x and y of $f(x, y)$). G stands for the 2-D Gaussian function:

$$G(x,y) = e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3.2}$$

Where σ is the standard deviation of the distribution.[18] Each of these components serve to address one of the two requirements for detecting intensity changes in I, as can be illustrated by applying the operator to a gray-scale natural image.

First, G allows the filter to fulfill the requirement of operating at different spatial frequency scales since it serves to blur an image based on the value of σ: any features of the image at a spatial frequency less than σ are washed out. Thus, a smaller the value of σ blurs out all but the most high-frequency features, while a larger value of σ blurs out all but the coarsest spatial features. For example, Figure 3.4 depicts an image that has been filtered for different levels of σ. In this respect,

[18] When the exponent of G is given an appropriate base, it becomes a *probability density* function.

G operates as a low-pass filter relative to the value of σ.[19] As a filter G also has the virtue of being smooth and localized both in terms of spatial position and spatial frequency.

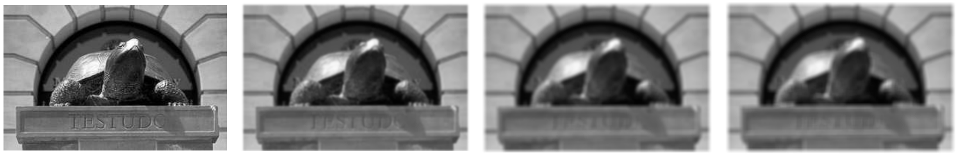


Figure 3.4: Application of Gaussian Filter.

Second, ∇^2 is a differential operator, which was the other requirement on an adequate filter, and it is also orientation independent. The virtue of the orientation independence is that it allows the filter to detect a zero-crossing regardless of its orientation in I. When convolved with G, one can see that $\nabla^2 G$ has a "Mexican-hat" shape as depicted in Figure 3.5. Figure 3.5 also depicts the operator in 2D, with gray values indicating 0 values in the matrix, white values indicating positive values, and dark gray indicating negative values. Applying a $\nabla^2 G$ mask to a natural image, for different value of σ (as in Figure 3.6), does indeed reveal many apparent zero-crossings, that is, intensity changes, in the image.[20]

[19] In image processing, a Gaussian mask is applied to each individual pixel. The mask is a matrix which approximates the form of the 2-D Gaussian distribution for some σ. When the mask is applied to the pixel, the value of the pixel is determined by taking a weighted average of the values in the matrix. The mask is centered on the target pixel, and that value receives the greatest weight, with low values at the edge of the matrix. Changing σ changes the "spread" and hence the relative weight of each point in the mask.

[20] Note the filtered images in Figure 3.6 only show the zero-crossings detected by $\nabla^2 G$ at different spatial scales.

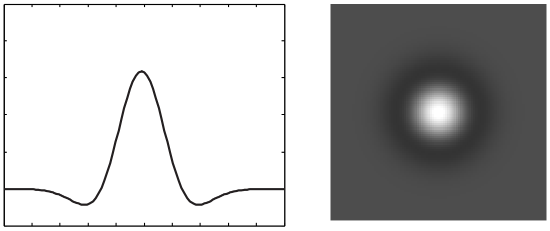


Figure 3.5: The Shape of the $\nabla^2 G$ Filter.

Marr and Hildreth hypothesized, based on then contemporary results in psychophysics and neurophysiology, that the retina contains channels with receptive fields that approximate the form of $\nabla^2 G$; in effect, the retina contains an array of $\nabla^2 G$ filters at different scales. If one inverts $\nabla^2 G$ (an upside down hat), then they reasoned that adjacent retinal cells with overlapping ON-center (hat up) and OFF-center (hat down) receptive fields would serve to detected a zero-crossing.[21] A layer of channels with overlapping receptive fields, and different space constants (i.e., values of σ), would be the analog of the multiple filtered versions of the image in Figure 3.6.

The first part of Marr and Hildreth's theory provides an account of the operation carried out by the retina in order to detect raw intensity changes in the retinal

[21] Specifically, they had in mind certain classes of retinal ganglion cells (RGC) which are the "output" cells from the retina. The receptive fields of the most well-known RGCs are known to have a characteristic "center-surround" organization, so that whether a RGC fires is based on the level of light intensity in the center or surrounding ring of the field. A RGC with a center-ON receptive field only fires when light is present in the center region (or when there is no light present in the surround region), while RGCs with center-OFF receptive fields have the reverse firing behavior.

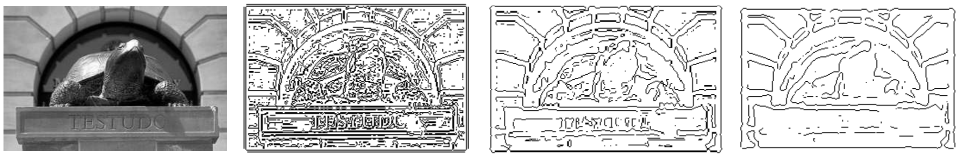


Figure 3.6: Application of $\nabla^2 G$ Filter.

image. Thus, $\nabla^2 G$ forms part of the what-component of their computational theory. However, this first part of the theory does not yet warrant appeal to primitives of a representational scheme, such as an edges or bars. In describing the first part of Marr and Hildreth's theory, Marr (1982, p.68 emphasis in original) notes:

> Up to now I have studiously avoided using the word *edge*, preferring instead to discuss the detection of intensity changes and their representation by using using oriented zero-crossing segments. The reason for this is that the term *edge* has a partly physical meaning—it makes us think of a real physical boundary, for example—and all we have discussed so far are the zero-values of a set of roughly band-pass second-order derivative filters. We have no right to call these edges, or, if we do have a right, then we must say so and why.

The distinction between an physical edge and a light intensity "edge" is central to theorizing about vision: "because the true heart of visual perception is the inference from the structure of an image about the structure of the real world outside" (ibid). Thus there is the issue of how to go from the detection of intensity changes to the representational primitives (edge, blob, etc.) of the raw primal sketch. The main difficulty is reflected in Figure 3.6 above: how and why does the visual system combine the information from $\nabla^2 G$ filters set to different scales? There is no a priori reason why the visual system should combine information about zero-crossings detected at different scales. However, there is a an external *physical* reason for combining the outputs of the equivalent channels in the retina. Marr and Hildreth took it as a further constraint that the physical phenomena that cause intensity changes are spatially, and hence, retinotopically localized—indeed, this was part of the motivation for the requirement of an image filter that is smooth in both the spatial frequency and spatial location domains, as is the case with G. If a physical phenomenon produces an intensity change at one scale, it is likely to produce a change in the same position, in the same orientation, at other spatial frequency scales as well. Thus, Marr and Hildreth (1980, p.202) theorized that the organization of the visual system respects an important assumption about the source of zero-crossing segments detected in I:

> If a zero-crossing is present in a set of independent $\nabla^2 G$ channels over a contiguous range of sizes and the segment has the same position and orientation in each channel, then the set of such zero-crossing segments may be taken to indicate the presence of an intensity change in the image that is due to a single physical phenomenon (a change in reflectance, illumination, dept, or surface orientation).

Marr and Hildreth called this the *spatial coincidence assumption.* According to the assumption, if a zero-crossing is detected by a high-spatial frequency $\nabla^2 G$ channel and a low-spatial frequency $\nabla^2 G$ channel, then they can be taken together.

It is by combining the outputs of $\nabla^2 G$ channels in this way that Marr and Hildreth hypothesized that the raw primal sketch is constructed, with labels for edges, blobs, and bars assigned to groupings of zero-crossing segments localized to regions of I. It is the raw primal sketch that constitutes the first representational scheme for which the primitives: "have a high probability of reflecting physical reality directly" (Marr, 1982, p.71).

Marr and Hildreth's theory has had a greater influence on image processing than on research on the function of the retina. For example, $\nabla^2 G$ is a staple filter in image processing[22], but there is no evidence that the receptive fields in neurons in the retina approximate $\nabla^2 G$.[23] Still, the theory provides an elegant illustration of Marr's information-processing framework. The information-processing task to which the raw primal sketch is suited is the detection and representation of intensity changes in I. The process in question is the mapping from the retinal image I to the raw primal sketch. What is the computational theory of the retina for the task of detecting and representing intensity changes? Roughly, the what-component of the computational theory is the combining of outputs from different $\nabla^2 G$ channels in order to token different primitives for describing (i.e., representing) the intensity changes in I. The why-component is provided by the various constraints identified by Marr and Hildreth for the task of detecting and representing intensity changes:

[22] Though other filters—notably, that of Canny (1986)—are better at detecting edges in natural images.

[23] Specifically, the form of center-surround receptive fields of RGCs do not approximate the form of $\nabla^2 G$. For some discussion, see citetTroy2002.

(i) the filter for I must operate at different scales and be smooth; (ii) it must be a second-order derivative operator (since detecting intensity changes is operationalized as the task of detecting zero-crossings); and (iii) the combining of the outputs of the filter channels must respect the fact that physical phenomena that produce intensity changes are spatially localized. (i) and (ii) almost uniquely identify $\nabla^2 G$ as the appropriate operator for the purpose of detecting intensity changes, and (iii) motivates the spatial coincidence assumption. Marr and Hildreth's theory was also suggestive of the neural implementation (in the retina), and the nature of the representational schemes (I and the raw primal sketch).

While many of Marr's specific theoretical proposals have perhaps not aged well, his approach has had a lasting influence on how researchers think about vision within the information-processing framework. I now turn to discussion of the content-commitments of Marr's approach.

3.5 The Content-Component of the Level 1 Analysis

I think that there can be little doubt that Marr's approach appeals to *some* notion of representation with accompanying content. Marr is explicit that his framework rests on the duality between representation and process, and part of how a representational scheme is defined is by reference to what it represents; namely, the information it makes explicit. A combination of tokened primitives from a representational scheme also serve to name, or stand in for, something in the target domain of the representational scheme. For example, the primitives of the raw primal sketch function to make explicit information about intensity changes in the retinal image.

In order to make my argument for my first conclusion, three issues need to be addressed: (1) if a notion of content is so crucial to his approach, then it should have a clear place in Marr's explanatory hierarchy; (2) we need to see more of an argument for why the content is original; and (3) we need an argument for why having the ingredients for internal representation are indispensable to the role of representational schemes in Marr's explanations of visual phenomena. In this section, I take up each of these tasks in turn.

3.5.1 The Place of the Component

If content is so crucial to Marr's approach, then we should be able to place it at one of his three levels of analysis. The most appropriate position, I believe, is to place it at the top. This comports with many other approaches to explanation in cognitive science that refer to a "semantic" (Pylyshyn, 1984), or "knowledge" (Newell, 1982) level as the top-most level of analysis. It is also common interpretation of Marr to associate representational content with his Level 1 analysis (Burge, 1986; Peacocke, 1994).[24] However this interpretation is usually paired with two unfortunate claims.

First, many have presumed that the content in question must be *intentional*

[24] Interestingly, this is the same interpretation adopted by some proponents of Bayesian approaches to perception and cognition (Griffiths et al., 2010). Roughly, the idea is that Bayesian models characterize at Level 1 what is represented, and what inferences are made, but without commitments regarding the other two levels. This connection to Bayesian modeling is discussed by Bechtel and Shagrir (2015).

(Burge, 1986). Second, since Marr appears to treat the computational theory as coextensive with Level 1 in his hierarchy of explanation, this has resulted in many claiming that some notion of content is *part* of the computational theory. Here is how Shagrir (2010, p.494) describes this interpretation:

> Most philosophers take it that the role of all computational-level theories is to provide an information-processing description of the task, in terms of the visual (representational) content of neural states.

This identification of the computational theory with representational content has resulted in the bizarre assertion that Marr is somehow "wrong" to talk about Level 1 as involving a "computational" theory (see e.g., Ramsey, 2007, p.41 n.3; Sterelny, 1990, p.46). However, such a view is itself surely mistaken, given how clearly Marr articulates his idea of a computational theory (Shagrir, 2010).

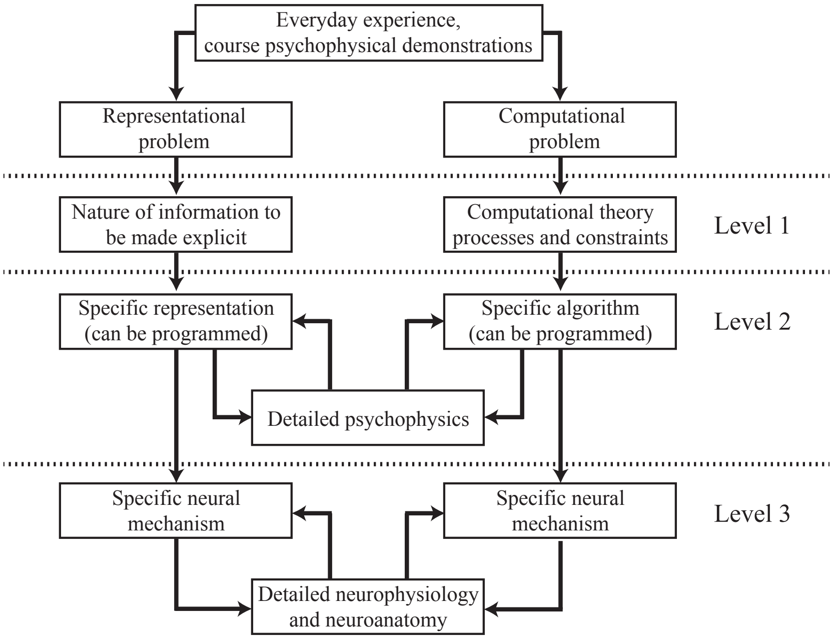
Fortunately one can maintain that representational content is a component of the Level 1 analysis, while also avoiding a commitment to these two claims. First, I do not think we need to (yet) burden Marr's notion of representational scheme with a commitment to intentional content as some have been want to do. And I will return to this issue below when discussing the Job Description Challenge. Second, I think that we should reject the common interpretative assumption that the computational theory is co-extensive with the Level 1 analysis. Instead, I believe a good case can be made that beyond the what- and why-components of the computational theory the representational content—the domain of entities or information made explicit by a representational scheme—provides another distinct component, what I will call the *content*-component of Level 1 explanation. There is certainly a position for this component to occupy, for Marr clearly intended the duality between representation and process to manifest itself at all three levels of explanation. And I think it is plausible that for the representational part of this duality, the content of a representational scheme—what information it makes explicit—finds its station at Level 1. Textual support for this interpretation can readily be found in Marr (1982).

First, recall that for Marr a process is always from one representational scheme to another representational scheme, and hence from one kind of information to another kind of information:

> At one extreme, the top level, is the abstract computational theory of the device, in which the performance of the device is characterized as a mapping from *one kind of information to another*, the abstract properties of this mapping are defined precisely, and its appropriateness and adequacy for the task at hand are demonstrated. (Marr, 1982, p.24 my emphasis)

In this passage Marr clearly associates *what* is represented—that is, the information that is being made explicit—as being part of the Level 1 analysis of an information-processing system. If we allow that the computational theory is not supposed to be co-extensive with the Level 1 analysis, then including a content-component to the level coheres with the description that Marr provides.

Second, the inclusion of a content-component to Level 1 is justified on graphical



Everyday experience,
course psychophysical demonstrations
Representational problem
Computational problem
Nature of information to be made explicit
Computational theory processes and constraints
Level 1
Specific representation (can be programmed)
Specific algorithm (can be programmed)
Level 2
Detailed psychophysics
Specific neural mechanism
Specific neural mechanism
Level 3
Detailed neurophysiology and neuroanatomy

Figure 3.7: Relationships between representations and processes.

grounds. Figure 3.7 is a reproduction of Figure 6-1 in Marr (1982), with the following modification: I have added hashed lines to distinguish where I think his three levels map onto the diagram (I have also used the same label for my figure). Marr intended the figure to summarize in broad outline how evidence in vision science relates to his three level approach, and to depict the "duality" between representation and process. Marr introduces his three levels within the context of explicating the notion of a process, and in so doing makes reference to the representational scheme at Level 2, and its implementation at Level 3. It makes clear sense to include the representational content—the domain of entities, or information that is made explicit—at Level 1 as illustrated by this figure, which positions it at the same stage as the computational theory.

This interpretation, which is supported by the text, positions representational content at the appropriate level in Marr's framework, and does so without requiring a mischaracterization of the computational theory. It also appears (literally) to be what Marr had in mind.

3.5.2 The Originality of the Component

As mentioned earlier it is clear that Marr attributes a stand-in function to representational schemes. Although he does not say so overtly, I think it is straightforward why the content-component is original, based on how he defines the notion of representational scheme, and applies it to vision.

Recall that for Marr (1982, p.21) representational schemes serve to "capture some aspect of reality by making a description of it using a symbol". With respect to vision the aspects of reality are ones that are accessible via the information latent in the retinal image. And the purpose of each stage of visual processing is specified by what information in the retina it make explicit (Marr, 1982, p.37). So it is constitutive of the information-processing tasks (whether service tasks or the chief task) that they have "informational" content of some kind. And there is no reason to think this content is derived, since it is the result of the architecture of the visual system, not the intentions or designs of some agent.

For example, as we have seen the function of the (raw) primal sketch is to make explicit information about intensity changes, and we saw Marr and Hildreth's proposal for how this is done. According to their theory, it is able to represent this information because it is organized in a fashion that meets certain constraints.

The raw primal sketch constitutes a representational scheme for information about intensity changes because of how it is built, and in that sense it is original. To claim otherwise is, I believe, to radically mischaracterize both Marr's philosophical approach, and its application in explaining visual phenomena.

3.5.3 The Indispensability of the Component

I think it is also straightforward to see why the content-component, and hence a notion of internal representation, is so indispensable to Marr's framework. The simple reason is that if the phenomenon of interest is some information-processing task carried out by a device, then one's explanation of the phenomenon must include positing a content-component. This is clear from considering both Marr's example of the cash register as well as Marr and Hildreth's theory of edge detection.

Consider the cash register first. The initial information is of the price for each individual, and the information that the device is supposed to provide is that of the money owed by the consumer. The information-processing task of the device is to perform the transformation from the price of the items to the sum that is owed. Notice that further constraints and conventions are important to understanding the device. It is only because transactions are made by exchanging *currency* that the cash register is adequate to the task, and the information that is made explicit is prices and money-owed. Consider what we would lose in our description of the cash register, if we did not acknowledge the content of the input and output strings. What the numerals make explicit is not simply numbers, but *prices* in some currency. The constraints on the operation, such that addition is the appropriate function for the cash register to perform, rest on the mapping to the price of the costumer's items. That the device performs addition of prices, as opposed to some other function, depends on how we interpret the inputs and outputs of the device.

The argument applies equally well in the case of Marr and Hildreth's theory. The theory is supposed to explain a mapping from a representational scheme for light intensity, the retinal image I, to one that represents intensity changes in I, the raw primal sketch. What tokenings of primitives in the raw primal sketch represent are intensity changes in I; that is, the content-component for the raw primal sketch is information about intensity changes in I—it is what the representational scheme makes explicit about the light intensity values in the retinal image. The constraints identified by Marr and Hildreth are determined by the information-processing task in question, the detection and representation of light intensity changes. To remove the content-component from the explanation would undermine the justification that the what-component receives from the why-component of the theory. And as we have seen, Marr clearly attributes a stand-in function to his representational schemes, and unlike the cash register, there is no reason to think the content-component of these schemes is derived.

So the argument for the indispensability of internal representations to information-processing explanations (as Marr characterized them) is straightforward:

(M1) An explanation of visual phenomena is an information-processing explanation only if it posits representational schemes.

> (M2) A formal scheme is a representational scheme only if (i) its internal states serve a stand-in function within the visual system and (ii) have content that is original.
>
> (M3) Therefore, an explanation is an information-processing explanation only if it posits internal representations. [from (M2) and my recipe for internal representation]

(M1) reflects how Marr characterizes the core of his information-processing framework, as reflecting the duality between representation and process. (M2) follows, I believe, from how Marr defines a representational scheme. First, he explicitly attributes a stand-in function to representational schemes. Second, although a cash register might have derived content, there is no reason to think that the information made explicit by the representational schemes of the visual system are anything other than original. In which case, we can infer (M3), which entails my first main conclusion: a notion of internal representation is indispensable to the information-processing framework (under the proviso that Marr's approach can be taken as representative of the framework in general).

3.6 The Framework and the Anti-Representationalist Challenges

So far I have provided a summary of the core features of Marr's approach, which I illustrated with one of his theories. I further argued that a content-component is located at the top-most level of the approach, and that his notion of a representational scheme has all the ingredients for being a formal system of internal representation.

One of my reasons for discussing Marr's work is that it has been used to develop and support some of the anti-representationalist challenges I introduced in the previous chapter. According to these arguments, reflection on Marr's approach, and its application to vision, *shows* that no notion of representation plays an indispensable role in the information-processing framework. If these arguments succeed, then they undermine my first two main conclusions. However I believe these arguments rest on misunderstandings of Marr's ideas, at least when it comes to the notion of internal representation.

3.6.1 Informality and the Content-Component

According to Chomsky (1995, 2000), one does not find any notion of representational content in Marr's information-processing approach to vision:

> [The] work is mostly concerned with operations carried out by the retina; loosely put, the mapping of retinal images to the visual cortex. Marr's famous three levels of analysis—computational, algorithmic, and implementation—have to do with the ways of construing such mappings ... the theory applies to a brain in a vat exactly as it does to a person seeing an object in motion. (Chomsky, 1995, p.52)

For Chomsky, any discussion of Marr's that references representational schemes does not bring with it a commitment to a notion of representational content:

> No notion like "content", or "representation of", figures within the theory

> . . . when Marr writes that he is studying vision as “a mapping from one representation to another, and in the case of human vision, the initial representation is in no doubt—it consists of arrays of image intensity values as detected by the photoreceptors in the retina” (Marr, 1982, p.310)—where “representation” is not to be understood relationally, as “representation of”. (Chomsky, 1995, p.52-53)

The reason that Marr’s use of ‘representation’ is not to be understood relationally, according to the Informality Challenge, is that intentional descriptions within Marr’s approach are purely informal. For example, the visual system no more represents the distal environment than a comet “aims” at the Earth (to use Chomsky’s example). While I think the argument from informality likely succeeds when it comes to some explanations in cognitive science, the argument fails when motivated by Marr’s framework.

As I argued in the previous section, a notion of internal representation is indeed indispensable to the information-processing framework—conclusion (M3) from the last section. To deny this conclusion requires denying one of the premises of my argument. Both premises are well supported by Marr’s own writing: it is hard to deny the centrality of the notion of a representational scheme to Marr’s framework, and with a representational scheme comes the ingredients for internal representation. The notion of a representational scheme, and its content-component are in no obvious way a reflection of informal description. Consider again Marr’s work on edge detection. The retinal image realizes a representational scheme for light intensity, and the raw primal sketch provides a representational scheme for changes in light intensity. In both cases there is a clear sense in which they are understood to have representational content: it is the information that these schemes make explicit. The content-components in the explanation reflect the description of the information-processing task: detecting and *representing* intensity changes. Nothing equivalent to a content-component applies when describing a comet as aiming at the Earth.

There is no reason to interpret attributions of content in Marr's framework as being informal, akin to the use of intentional idioms in other branches of science. Thus so far my argument stands. Interestingly, Chomsky cites Egan (1995) as providing an accurate account of Marr's framework. Thus, I now turn to looking at how Egan has used Marr's framework to motivate the Instrumental Challenge.

3.6.2 Instrumental Content and the Computational Theory

Egan has relied extensively on Marr's work—in particular, Marr and Hildreth's theory of edge detection—in developing what I have called the Instrumental Challenge. I believe that Marr's work provides little support for her challenge, and hence, I do not it provides good grounds for rejecting my argument.

To start, let me distinguish between two issues. When it comes to computational models, there is indeed an issue of how they are to be interpreted in relation to a phenomenon of interest. A computational model at an abstract level, is simply a formalism. For example, at an abstract level $\nabla^2 G$ is an operator, combined by convolving ∇^2 and G. As it happens, a filter that approximates $\nabla^2 G$ meets certain constraints for a good image filter. Thus, when employed in the computational theory for an edge detecting device (the retina) the operator is interpreted in a certain way—as characterizing the form of receptive fields. The general point is that for any formal model, one must distinguish the model from its interpretation under some application. But a *separate* issue is whether a system to which a computational model is applied contains internal states that satisfy the recipe for some notion of representation. The *model*, not the system to which it is applied, has instrumental "content" because of how we interpret it. I believe that Egan's argument rests on a conflation of these two issues: the issue of (i) how computational models are interpreted with (ii) the issue of whether a system to which a computational model is applied is representational in some sense.

The gist of Egan's challenge is that the core of an information-processing explanation is the what-component of the computational theory, and it is only when one interprets the mathematical function specified by the what-component, in line with one's explanatory interests, that the mathematical function provides an explanation of the phenomenon in question. Thus, what "content" we attribute to the internal states posited by an explanation are always relative to explanatory interests, and fail to have my ingredients for internal representation (i.e. the content is derived). The main example used by Egan to maker her case is Marr and Hildreth's theory. (Egan, 1992, p.454) claims that Marr's computational theory is: "a *formal* characterization of the function(s) computed by various processing modules." Thus according to Egan, the computational theory consists solely of a specification of a mathematical function (c.f. Egan, 1995, p.187; Egan, 2010, p.255. Egan takes the use of $\nabla^2 G$ in Marr and Hildreth's theory to provide support for her view. In particular, Egan regularly cites the following passage from Marr in support of her idea that the computational theory is just a specification of a mathematical function.

> I have argued that from a computational point of view [the retina] signals $\nabla^2 * G$ . . . From a computational point of view, this is a precise characterization of what the retinal does. Of course, it does a lot more—it transduces the light, allows for a huge dynamic range, has a fovea with interesting characteristics, can be moved around, and so forth. What you accept as a reasonable description of what the retina does depends on your point of view. I personally accept $\nabla^2 G$ as an adequate description, though I take an unashamedly information-processing point of view.(Marr, 1982, p.337)

The conclusion that Egan draws from this passage is that the operator $\nabla^2 G$ suffices for specifying the computational theory for edge detection. The only way to connect the operator to the phenomenon of interest is by providing an interpretation of the operator that connects it to detecting changes in light intensity. Egan (1992, p.455) correctly observers that Marr avoids claiming that the primitives of the raw primal sketch (e.g. "edges" and "bars") represent properties of distal objects. But what she appears to also deny is that the primitives are intended to represent properties of intensity changes in I:

> When the computational characterization is accompanied by an appro-

> priate cognitive interpretation, in terms of distal objects and properties, we can see how the mechanism that computes a certain mathematical function can, in a particular context, sub-serve a cognitive function such as vision . . . . So when the input states of the [$\nabla^2 G$] filter are described as representing *light intensities* and the output states *changes of light intensity* over the image, we can see how this mechanism enables the subject to detect significant boundaries in the scene. (Egan, 2010, p.256 emphasis in original)

Egan is right that the what-component of a computational theory, in isolation from the why-component, might be characterized in purely formal terms, and so has no intrinsic connection to the phenomenon of interest. In which case, an interpretation of the what-component is required. But as I pointed out above, how we interpret a computational model of a system is a separate issue from whether the system is representational. For example, Marr and Hildreth propose $\nabla^2 G$ as a filter because it meets certain constraints—the why-component of their computational theory. It is true that it is only when we interpret $\nabla^2 G$ in a certain way (as characterizing the form of receptive fields of retinal cells) that we connect the operator to vision, and see how the operator satisfies the constraints. However, the constraints identified by Marr and Hildreth are *themselves* dependent on how the information-processing task is characterized. Marr and Hildreth's characterization includes a content-component: the raw primal sketch represents intensity changes in I. Since the identification of $\nabla^2 G$ as an ideal filter for detecting intensity changes is a result of the filter satisfying the appropriate constraints, and the selection of the constraints is dependent on how the information-processing task is characterized, the interpretation of the operator is also dependent on how the task is characterized. And it is the characterization of the information-processing task that in turn determines the content-component. Thus, it is because of the content-component that we are able to make sense of the constraints that make $\nabla^2 G$ the best filter for detecting intensity changes. To summarize my point in Egan's own terminology, the "pragmatically motivated gloss" of $\nabla^2 G$ is a reflection of the content-component of the computational theory. Therefore, it makes little sense to claim that the gloss is itself the content-component as Egan seems to suggest. [25]

The $\nabla^2 G$ filter, one part of Marr and Hildreth's theory of edge detection, is the primary example that Egan has used to develop the argument from instrumental content. However, her notion of instrumental content does not match the notion of content found in Marr's framework, and seems to mischaracterize the significance of the computational theory, and representational content, to his framework. In so far as Egan's challenge rests on her characterization of Marr's framework, it fails to undermine my argument.

[25] Thus, by claiming that the information-processing task for edge detection is the computation of $\nabla^2 G$—as opposed to detecting intensity changes in I—Egan (1999, p.192) appears to misunderstand the significance of the operator to Marr and Hildreth's theory.

3.6.3 Job Descriptions and Representational Schemes

Ramsey's Job Description Challenge applies to any notion of internal representation employed in cognitive science: in order for an internal state to be a mental representation, it must have a representational function tied to the state having intentional content. We further saw that intentional content requires ingredients of perceptual objectivity and robustness, at least when it comes to perceptual representations. So does Marr's notion of a representational scheme meet Ramsey's challenge? Burge (1986, p.35) seems to think so:

> the claim that [Marr's work on vision] is intentional is sufficiently evident. The top levels of the theory are explicitly formulated in intentional terms. And their method of explanation is to show how the problem of arriving at certain veridical representations is solved.

I agree with Burge that the Level 1 explanation includes a content-component, as we have seen. But I believe that the correct answer to the question is that being a representational scheme does not *simpliciter* satisfy the ingredients for being a mental representation. For example, some of the representational schemes that Marr posits in vision seem to fail the job description challenge, and merely describe sensory registers.

First, the retinal image I is a good example of a sensory register, which makes explicit information about the light intensity in the visual input. First, the content-component is original, and it serves a stand-in function, so it is an internal representation. Second, it is a sensory register because it is an internal representation of a sensory system that tracks (very simple) properties of the proximal stimulus. But it lacks the other ingredients for perceptual representation: its content is neither perceptually objective nor is it perceptually robust (though perhaps it is resilient).

Second, the raw primal sketch, which Marr (1982, p.71) describes as the first representational scheme with primitives that "have a high probability of reflecting physical reality directly", also appears to be a sensory register. The raw primal sketch represents intensity changes of different types (e.g., blobs, edges, and bars) of different orientations, scales, and locations, but these are also not properties of the distal world, but the retinal image itself (in this respect, the information that the raw primal sketch makes explicit may indeed be "internal"; see e.g., Segal, 1989). In a certain sense, the content of the raw primal sketch is perceptually robust, since it represents intensity changes, regardless of how they are caused (changes in illumination, orientation, or surface reflectance properties). But it does not represent these varied causes.[26]

So neither the retinal image nor the raw primal sketch are plausibly perceptual representations. However, Marr's own view about the content-components of different representational schemes in the visual system is that as one moves deeper into the visual system, subsequent representational schemes make explicit information about more distal (and hence objective) properties of the environment. For

[26] I think it is mistaken to suggest that tokening of primitives in the raw primal sketch represent a disjunction of distal properties, as suggested by Egan (1999, p.184). Marr states quite clearly that what the raw primal sketch represents is intensity changes, not a disjunction of their causes.

example, other simple organisms do not in fact represent distal properties at all, even though they represent in Marr's sense:

> When does a specific configuration in the image imply a specific configuration in the environment?. . . In a true sense, for example, the frog does not detect *flies*—it detects small, moving, black spots of about the right size. Similarly, the house fly does not really represent the visual world about it. . . We, on the other hand, very definitely do compute explicit properties of the real visible surfaces out there, and one interesting aspect of the evolution of visual systems is the gradual movement toward the difficult task of representing progressively more objective aspects of the visual world. (Marr, 1982, p.340)

It is clear from passages like this one that Marr believed that the human visual system represents distal properties of the visual world. Thus it seems he in fact endorsed the Distal Object Thesis. The notion of a representational scheme is broad enough to include such apparent examples of genuine perceptual representations, as well as cases of sensory registration. However, it is important to see that positing perceptual representations in this way not something that Marr *argued* for, but rather took for granted (Hoffman and Singh, 2012). This is make explicit by Marr himself, who took RTM as a background assumption to his approach:

> From a philosophical point of view, the approach that I describe is an extension of what have sometimes been called representational theories of mind. . . representational theories conceive of the mind as having access

> to systems of internal representations; mental states are characterized by asserting what the internal representations currently specify, and mental processes by how such internal representations are obtained and how they interact. . . This scheme affords a comfortable framework for our study of visual perception, and I am content to let it form the *point of departure* for our inquiry. (Marr, 1982, p.6 my emphasis)

So it appears Marr may have taken the line that since RTM is true, some internal representations are mental (cf. Pylyshyn, 1984). But we are not assuming RTM as a metaphysical thesis[27]. So if we want an argument for my second conclusion, which meets the Job Description Challenge, then we need to look elsewhere. In particular, we need an argument for why an internal representation that is posited to explain some visual phenomenon (in particular object recognition) must be perceptually objective and robust.

3.7 Conclusion

In this chapter I looked at Marr's philosophical approach, which in part spawned and popularized the information-processing framework in vision science. I argued that a notion of internal representation was indispensable to Marr's work, and in so far as it is representative of the framework at large, this argument amounted to a defense of my first main conclusion. I further showed that two

[27] It is also not even clear that assuming RTM is a plausible way to meet the Job Description Challenge (Ramsey, 2007, 38-67)

anti-representationalist challenges, which are motivated by Marr's work, fail. At the same time, Marr's work takes a notion of mental representation for granted, and so more must be done to provide an argument for Conclusion 2 that meets the Job Description Challenge. In particular, we need to see an argument for why having my mental ingredients of perceptual robustness and objectivity are indispensable to the role that an internal representation plays in the explanation of a visual phenomenon. I lay the groundwork for such an argument in the next chapter.

Chapter 4: Inferring Content from Constancy

4.1 Introduction

Let me recap. My ultimate goal is to show that some theories of content—in particular, informational theories—are relevant to some explanatory practices in vision science—in particular, those aimed at explaining facts about visual object recognition. My second way-point to this end, Conclusion 2, is to show that a notion of perceptual representation is indeed indispensable to these practices.

In Chapter 2, I laid out a recipe for perceptual representation: a list of ingredients for when an internal representation of a sensory system is a form of mental representation, as opposed to a mere register of proximal stimulation. I also canvased some arguments for why these ingredients are missing from the internal states posited in the explanations of visual phenomena. Focusing on Marr (1982), in Chapter 3 I looked in detail at the information-processing framework in vision science—which subsumes work on visual object recognition—in order to assess some of the anti-representationalist challenges from Chapter 2. These arguments, I claimed, do no justice to the structure of the framework. I also ended up with something of a disjunctive conclusion: that the representational schemes central to the framework could either be mere registers of sensory input or genuine perceptual representations.

Ultimately, whether the key ingredients for perceptual representation—namely, perceptual objectivity, and robustness—are present will depend on facts about the phenomenon that is being explained. I believe that the internal representations that underlie the human capacity for visually recognizing objects satisfy the two crucial ingredients required for genuine perceptual representation. In this chapter, I lay the groundwork for my argument to this conclusion.

Vision exhibits perceptual constancies: what we see appears relatively stable despite changes in proximal stimulation. Historically, the fact that vision exhibits perceptual constancies has been used as grounds for arguing that vision is perceptually objective (Boring, 1946; Brunswik, 1940; Cassirer, 1944; Thouless, 1931), and perhaps even representational (Cooper and Hochberg, 1994; Dretske, 1981). Indeed, recently Burge (2010) has made the same sort of argument in defending the indispensability of perceptual representations as a posit for perceptual psychology at large. I will call this argumentative strategy, which moves from facts about perceptual constancies to conclusions about perceptual objectivity and representation, the *argument from constancy*.

In this chapter I assess the prospects of the argument from constancy. I present three prima facie challenges for the argument, which I refer to as the *Dependence Challenges*. The upshot of my discussion of these challenges is that if an argument from constancy is to succeed, it should be narrowly focused on object constancy (or "invariance") in vision—which is the primary subject matter of research on visual object recognition. In contrast, Burge (2010) offers a very broad argument from constancy, which he runs for virtually all forms of perceptual constancy in vision. I argue that Burge does not have the resources for meeting the dependence challenges.

The rest of the chapter is structured as follows. In Subsection 2, I spell out the background of the argument from constancy, and how perceptual constancies feature in the story of vision. In Subsection 3, I provide a (re)construction of the argument from constancy using my recipe for perceptual representation. In Subsection 4, I lay out the Dependence Challenges, which suggest a need for a narrower argument from constancy focused on object constancy. In Subsection 5, I summarize the basis for the broad argument from constancy offered by Burge (2010). In Subsection 6, I argue that Burge's version of the argument falls prey to the Dependence Challenges. Finally, in Subsection 7, I conclude the chapter.

4.2 Perceptual Constancies and the Story of Vision

Let me say a bit more about what perceptual constancies are like. In general terms, a *perceptual constancy* is the fact that some form of perception remains unchanged in the presence of certain kinds of changes to the sensory input. The most familiar forms of perceptual constancy come from vision. To illustrate, consider again my brother's cat, Mr.Muscles, who I observe first sitting by the window, and then jumping to the floor and walking through a pool of sunlight. My perception of the cat on his trip across the room exhibits the main kinds of perceptual constancy found in vision:

> *Shape constancy*: The perceived shape of an object appears constant across changes in retinotopic shape. Example: I perceive the cat as

having the same overall shape, despite changes in the shape of his retinal projection as he moves across the room.

Size constancy: The perceived size of an object appears constant across changes in retinotopic size or perceived distance. Example: I perceive the cat as having the same size despite changes in distance as he moves from the window across the floor.

Color constancy: The perceived color of an object appears constant across changes in illumination. Example: I perceive the color of the cat's coat as unchanged despite variation in spectral illumination of his fur as he walks across the room into the pool of light.

Object constancy: The perception of an object as an object, and its identity, appears constant across changes of view-point (e.g., changes in retinotopic size and shape, illumination, orientation, and occlusions). Example: despite the many changes in the retinal input, and partial occlusion of his own body (e.g., of his legs while walking) I do not perceive a change in objecthood, or identify, as Mr.Muscles traverses the room.

Based on these examples, it is easy to see (literally) that virtually all of our visual experience of the world exhibit some kind of perceptual constancy. Historically, constancies have been central to how psychologists characterize perception in general, and vision in particular. They are fundamental to the story that vision science tells about the nature of vision, and the relationships between the distal world, proximal stimulation, and perception (Epstein, 1977; Ittelson, 1951). This role of constancies in perceptual theory provides the background and basis for the argument from constancy.

In this section, I first review a kind of main narrative in vision science regarding the relationship between the distal world and the proximal stimulus, and why whatever it is we perceive visually, it is not simply the proximal stimulus. Next, I outline how perceptual constancies have been central to whether one adopts what I will call "objective" or "subjective"' readings of this narrative. It is the objective reading of the narrative that is at the heart of the argument from constancy.

4.2.1 Some Problems in the Main Narrative

The extent to which we see the world as a coherent whole may seem like a platitude. However this mundane fact about our visual experience becomes remarkable when we consider the relationship between proximal visual inputs and the distal world. In particular, two facts about the mapping from the distal world to the proximal stimulation of the retina are typically characterized as "problems" that the visual system must overcome in order to afford us a coherent perception of the world. These problems also provide the traditional grounds for thinking that the objects of perception are distinct from the events of proximal retinal stimulation (Epstein, 1977).

The Underdetermination Problem. We are accustomed to seeing the world as a single unambiguous whole. However, for any given retinal stimulation, there are infinitely many possible states of the distal world that could have caused the very same stimulation. This is known as the *underdetermination problem*, which can be illustrated using the perception of size and color.

Consider a retinal projection of a simple 3D cube at some distance from the viewer. Isolated from the inputs to the rest of the visual field, there are infinitely many possible states of affairs in the distal world that could have caused the very same retinal projection: squares of different sizes at different distances, but also parallelograms at different orientations, also varying in size and viewing distance, and with appropriate variation in light source (Kersten and Yuille, 2003). Thus the retinal projection is underdetermining of the size (and shape) of the distal object that caused the retinal projection. Similarly, consider that with three cone receptors in the human retina, the spectral distribution of light that reaches the eye is reduced to three values, even though a spectrum requires an infinite number of values to specify its energy at each wavelength. This has the consequence that there are an infinite number of spectra that, given our trichromatic vision, are treated as equivalent. Such equivalence is illustrated by the phenomenon of metamers: distinct spectral reflectance distributions that are perceived as being equivalent.

Although we are normally unaware of the underdetermining quality of input to the eyes it is readily revealed by visual illusions, and multi-stable figures such as the Necker Cube and Face-Vase (Figure 4.1), in which the visual system can alternate between multiple percepts, or “interpretations”. In the case of the Necker Cube, we see the circled vertex as either in the foreground or background, resulting in a change in the perceived 3D orientation of the cube. In the case of the Face-

Vase, we can see either the black or white portions of the image as the figure or ground, resulting in a change in percept between two symmetrical faces in profile, or a centrally positioned vase.

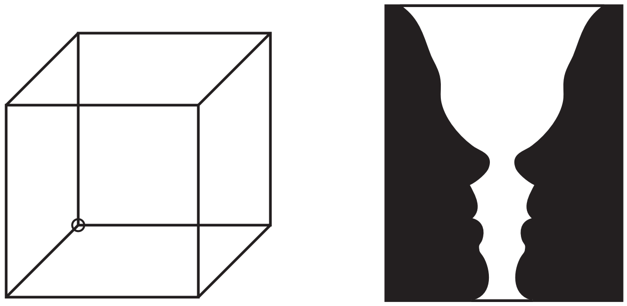


Figure 4.1: The Necker Cube and Face-Vase.

What does the underdetermination problem teach us about visual perception? One textbook lesson is that the object of perception has greater *specificity* than what can be extracted from the proximal stimulus alone. So the object of perception is not a property of the incidental proximal stimulation itself. Instead our visual system constructs a percept using more information than is made available just in the retinal input. The existence of visual illusions and multi-stable figures tells us that some kind of *process* is required, as Marr asserted. For example, it has been suggested that visual processing involves "taking-into-account" separate information not available in the retinal image (Epstein, 1973); requires some kind of "unconscious inference", as Helmholtz and many others have believed (e.g., Rock, 1983; for a historically oriented review, see Hatfield, 2002); or depends on the internalization of environmental constraints (Marr, 1982; Shepard, 1984). While each of these ideas remains fashionable today (in one form or another), Bayesian approaches to vision have also provided useful abstract characterizations of visual processing in terms of probabilistic inference (Kersten and Yuille, 2003). Despite differing in the details, all of these approaches have in common the objective of explaining how the visual system constructs a percept, or rather, constructs a representation, that goes beyond the underdetermining information found in the retinal image.[1]

The Variation Problem. Just as we are accustomed to seeing the world unambiguously, we are also used to seeing the world as stable. However, no distal event or object ever produces the same retinal input twice due to differences in viewing conditions. Call this the *variation problem.* Consider again the examples of perceived size and color. As mentioned earlier, perceived size is generally constant across changes in viewing distance, while perceived color is constant with respect to changes in surface illumination. However, there are infinitely many possible distances at which we might view an object, resulting in differences in the projected retinal size of an object. Likewise, illumination conditions are never quite the same when we view an object. Normally we are unaware of the inherent variability in retinal inputs; however, this property of our visual input becomes salient when one wears inversion goggles, which use prisms to invert the light entering the lenses so that wearers perceive the world as upside down. When worn for the first time, visual percepts can be incredibly disorienting, though with training and experience individuals can become quite adept at navigating the world while wearing the goggles.

[1] Not everyone believes that the underdetermination problem warrants the claim that vision is an inferential process. For a recent argument against this common view see Orlandi (2014).

Overcoming the variation problem requires the visual system to make generalizations and predictions across input conditions, which might reflect innate or learned aspects of visual processing. The problem also reveals why the objects of visual perception cannot be specific to the proximal input, since our perception of the world is constant despite the substantial changes in retinal stimulation. In general, all research on perceptual constancies in vision can be seen as attempts to explain how the visual system overcomes the variation problem with respect to different dimensions of visual input.[2]

These two problems constitute received facts in mainstream vision science regarding the mapping between the distal world and proximal stimulation. Providing an explanation of how the visual system overcomes these problems, as they arise for different aspects of visual perception, remains a primary objective for research within the information-processing framework (see e.g., Rust and Stocker, 2010).[3] Given that these problems exist, yet we see the world as we do—as an unambiguous and stable whole—tells us that the objects of visual perception do not reduce to the transient proximal inputs themselves. But what about the mapping between perception and the distal world? In answering this question, perceptual constancies have traditionally taken center stage.

[2] A fact also true of research on visual object recognition, as we shall see in the next Chapter.

[3] Notoriously, Gibson (1950, 1979) rejected the existence of both problems, endorsing instead a "one-to-one-to-one" mapping between the distal world, proximal stimulation, and perception. However even to some of his followers in ecological psychology, rejecting the existence of these problems is considered empirically untenable (Withagen and Chemero, 2009).

4.2.2 Objective and Subjective Readings

Traditionally there have been two kinds of "readings" of perceptual constancies with respect to where we should locate the objects of perception.[4] According to what I will call *objective* readings, perceptual constancies reveal the respects in which vision is indeed objective: the objects of perception are non-perspectival properties of objects and events in the distal world. According to what I will call *subjective* readings, the object of perception is not something distal, but a higher-order property extracted or constructed from some function of the retinal image. For example, a classic proposal is that, in line with principles of Euclidean geometry, if visual angle is known, then perceived size and perceived distance vary proportionately (Epstein et al., 1961; Kilpatrick and Ittelson, 1953). Were this generalization true, a possible subjective reading would be that perceived size is a function of visual angle and visual cues for perceived distance.

These two sorts of readings are in large part motivated by focusing on one of two phenomena related to perceptual constancies (Ittelson, 1951). The first is constancy *achievement*, the fact that we appear to have largely accurate perception of the world across transformations of the visual input. The second is the *mechanisms* that underlie constancies, which are operational both during apparently veridical and illusory perception.[5]

[4] The expression "object of perception" is common in early work on perceptual psychology. In plane terms, it is simply *what* we perceive.

[5] The use of "mechanism" here is from Ittelson (1951), and should not be taken to presume a form of mechanistic explanation in the sense familiar from philosophy of science (Machamer et al.,

Objective readings are generally motivated by reflection on the achievement of constancy. For example, in early work on size constancy, Brunswik (1940) was convinced of an objective reading by the fact that judgments of size correlated more with distal properties of his stimuli than perceived retinal size, and likewise Thouless (1931) believed his work on shape perception showed that vision involves a "regression to the real object". Here is the approximately contemporaneous philosopher, Cassirer (1944, p.34-35 emphasis in original), explicitly making the argument from the achievement of perceptual constancy, to an objective reading:

> In modern psychology it appears clearly that there exists a peculiar function to which perception owes its objectivity. The "true" shape, the "true" size of an object are by no means that which is given in any particular impression, nor need they be the "sum" of these impressions. . . The *constitutive factor*. . . manifests itself in the possibility of forming invariants. Owing to this possibility, there exist for us a "perspective of illumination" and a spatial perspective and thus the perception of "objective" reality. . . If there were no [perceptual] constancy, we would, as it were, abandon ourselves to every change in external conditions; it would be impossible to segregate "things" and "properties" from the stream of becoming. To use Heraclitus' metaphor, we should be unable to "step down twice into the same river". . . Thus [modern] psychology. . . dismisses the dogma of the strict one-to-one correspondence between physical stimuli and perceptions. It is, on the contrary, the "transformed" impression,

2000).

> i.e., the impression as modified with respect to various phenomena of constancy, which is regarded as the "true" impression, since we can on these grounds construct knowledge of reality.

The core of Cassirer's argument is that since it is the "real object" that remains constant across transformations of the retinal input, it follows that the object of perception is something in the distal world. Hints of the same argument are present in the information-processing framework. For example, it is suggested by the very idea that vision is hierarchical: that at progressive stages of visual processing (and brain regions), there is increased invariance, and greater specificity of vision to the distal world (e.g., Rust and DiCarlo, 2010). Likewise, the same sort of reasoning is seemingly tacit in Marr's discussion of why internal representations of 3D shape, as opposed to earlier stages such as the 2D sketch, are directed at distal properties of the world.

In contrast, subjective readings have typically been motivated by focusing on the mechanisms of perceptual constancy. Illusions are of interest to vision scientists because they afford a means of investigating the mechanisms that appear to afford accurate perception of the world under normal viewing conditions (Coren and Girgus, 1977; Gillam, 1998). However if the same mechanisms are employed for both "veridical" and illusory perception, then one can argue that the same objects of perception are present under both conditions. Along these lines, here is Epstein (1977, p.7) defending a subjective reading:

> Most investigations of the constancies are interested more in constancy

> mechanisms or processes than in constancy as product. Nowhere is this more evident than in the frequent attempts to confirm hypotheses about the basis of constancy through experiments on illusions. The general opinion is that although constancy and veridicality may be distinguished from illusion on the basis of extra-perceptual criteria, that is, conventional assessments of correspondence between reports and physical measurement, veridicality and illusion differ neither in the character of the experience nor in the underlying process. . . Thus it is plain that it is constancy and not veridicality that is the focus of concern and it is the mechanism and not the achievement that commands interest.

Since the object of perception is that which remains constant with respect to sensory transformations, and there is nothing to distinguish—by perception alone—between veridical and illusory percepts, it follows that whatever the object of perception is, it is some function of the sensory signal, and not a property of the distal world. This subjective reading can be found historically in Gestalt Psychology (Koffka, 1935), but also in the information-processing framework as well. Recent work within the framework rebelling against objectivity in perception tend to focus on the same sort of reasoning as Epstein (e.g., Mark et al., 2010).

How then, are we to decide between the two readings? Needless to say, vision science provides no universally agreed on answer. Both readings are motivated by facts that cannot be discounted (Ittelson, 1951), so matters will depend (among other things) on which offers the best explanation of particular constancies in vision.

What I wish to now spell out is how an objective reading underlies the argument from constancy.

4.3 (Re)constructing the Argument from Constancy

Not all research carried out within the information-processing framework in vision science is focused on explaining perceptual constancies (although a great deal is; Palmer, 1999). Nonetheless, constancies are still fundamental to how researchers think of the visual system. This is most salient, again, with respect to the very idea that vision is hierarchical, or that at successive stages of visual processing, the representational schemes transition from tracking properties of the retinal image, or sensory signal, to properties of objects in the distal world—the very narrative within the story of vision that we began with in Chapter 1. In fact, the sort of considerations regarding perceptual constancies that historically motivated an objective reading, have also been the basis for positing mental representation in the visual system. For example, Cooper and Hochberg (1994, pp.224-225 emphasis in original) make this very claim:

> Most of the purposive behavior of which we are aware is directed toward the still and moving objects around us, or at the layout of surfaces within which we move. We must assume that evolutionary endowment and individual learning provide an accuracy of perceiving and remembering the objects of our behaviors, that is, at least at the level required for survival. Thus, our reports of objects' surface reflectance and shapes are

> often in better correspondence with the objects in the world than one might expect from local measures of their retinal images. In experimental research, on the object constancies, for example, observers' reports of such attributes as object shape, size, and lightness remain relatively invariant over changes in the retinal image due (respectively) to changes in orientation or slant, distance, or illumination on the object's surfaces. *Indeed, the very conception of mental representations was introduced as an essential component of perceptual theory in order to explain how such object constancy can hold, despite changes in the sensory information impinging on the receptor surface.*

So arguably, perceptual constancies have *always* provided a primary basis (even if only tacitly) for positing perceptual representations in the visual system.[6] Despite the apparent historical connection between perceptual constancies and perceptual representation, lacking is an explicit statement of how an argument from the one to the other is supposed to go. In this section I try to address this shortcoming by providing an explicit reconstruction of the argument from constancy.

[6] Although the paper this passage comes from is not well-known, I consider the opinion expressed to be rather illuminating, given the historical bend, and significance, of Hochberg's work. For example, he organized an early symposium (Hochberg, 1957) on the influence of the Gestalt movement in psychology (at which Brunswik, Gibson, and Kohler were participants); and edited a retrospective volume on experimental psychology in the 20th Century (Hochberg, 1998). Some of his work (spanning five decades) is collected in Peterson et al. (2006).

In the last chapter we left off with a disjunctive conclusion: either the representational schemes of the visual system are collections of sensory registers or perceptual representations. The missing special ingredients for selecting the second disjunct were perceptual objectivity and perceptual robustness. I believe an argument can be constructed from facts about perceptual constancies, and aspects of the story of vision, to the claim that a system of internal representations that explains these constancies must have these special ingredients. In this section, I first spell out more fully why perceptual constancies might allow an inference to perceptual objectivity. Next, I connect the problems faced by the visual system (the underdetermination and variation problems) to perceptual robustness. I then provide a breakdown of the structure of the argument from constancy.

4.3.1 A Boring Principle and its Application

Underlying both the objective and subjective readings is a kind of principle that I would like to make overt. As has been long observed, it is not mere change in stimulation that is important to thinking about perceptual constancies, rather it is *continuous change* (Ittelson, 1951; Koffka, 1935). Furthermore, it is because they are often continuous that constancies are typically characterized as reflecting a kind of *invariance.*

The notion of invariance comes from mathematics, where an invariant is a feature of a group or class of mathematical objects, which remains unchanged under certain transformations. For example, the angles of a triangle are invariant with respect to transformations of scale or rotation in a geometric space. Since perceptual constancies came to prominence in thinking about the nature of perception, it has been common to understand them as a kind of invariance: stability in perception in the face of transformations of the input.[7] What does the notion of invariance get us beyond the intuitive idea of a constancy? In part, the import of the notion was that it helped to focus inquiry on the investigation of invariance in perception under continuous, or parametric, *transformations* of stimuli. Introducing the notion was important because it provided a way of characterizing the sorts of generalizations that experimental psychology should hope to capture. Indeed, according to Stevens (1951), it was the investigation of invariance that made perceptual psychology a science.

Inspired by Dewey (1896), Gibson (1950), and Stevens (1951), it was Boring (1952) who first explicitly recognized the significance of the notion of invariance for discovering the objects of perception. In discussing size constancy, Boring reasoned as follows:

> If we could find a function, ϕ, that would be invariant when perceived size is invariant. . . then we could say even better what it is that is being perceived (invariant). In short, if perceived size is invariant when this function, ϕ, is invariant, then in judging size, you are perceiving not object size, not retinal size, but ϕ. To discover the object of perception, you have to discover what function of the parameters of the stimulus is

[7] Of course, the notion of invariance in perception is not the *same* as the mathematical notion, since the former is relative, while the latter is absolute. For discussion, see Cutting (1983).

> invariant when perception is invariant. (Boring, 1952, p.146)

For Boring, the object of perception is whatever is recoverable from the stimulus that remains invariant when perception is invariant. Let us call this the *Boring Principle*. It is clear that the principle is at work in the arguments we saw for both objective and subjective readings of perceptual constancies. For what is invariant, under the objective reading, is a property of something in the distal world, while under the subjective reading, it is a higher-order property extracted from proximal inputs. In both cases, we identify the specificity of the object with a function of the input that is invariant when perception is invariant. So crucially the Boring principle makes the object of perception contingent on what we discover to be an invariant function of the stimulus.

It is the Boring Principle that I wish to harness in laying out the argument from constancy. A first step to doing so is to consider how perceptual constancies are characterized by explanations within the information-processing framework. The Boring Principle tells us how to determine the relative specificity of what we perceive, based on invariance in perception. Thus, we must ask what is specific and invariant in perception from the perspective of representation and process in vision. Both are clearly tied to the contents of perception (cf. Yilmaz, 1967), in which case it is natural to consider them properties of representational content within the information-processing framework—or what I termed, in Chapter 3, the "content-component" of the Marr's computational level. In other words, when explaining specificity and invariance in perception, research within the information-processing framework posits an internal representation, which has content that exhibits the relevant levels of specificity and invariance.

If invariance and specificity are supposed to be explained as properties of representational content, then it is straightforward to see how we can use constancies to establish that the content of an internal representation is perceptually objective. Consider a representational scheme in some stage of processing in the visual system, which exhibits some level of invariance with respect to transformations visual inputs. In line with my argument from the last chapter, I will assume that many of the staple and special ingredients for perceptual representation (enumerated in Chapter 2) are satisfied by the representational content of the scheme: for example, the content is original (as opposed to derived), and internal states within the scheme functions as "stand-ins" of what they are about for the purpose of mental processing. If the content of some internal representation is invariant with respect to transformations that are sufficient for making the content distally specific—in line with an objective reading of the empirical facts about the constancy—then the content is non-perspectival. And since the notion of objectivity we are working with from Chapter 2 is that of being non-perspectival, then the content is perceptually objective. Or put differently, if the invariance that is supposed to be explained by a set of representations obtains with respect to viewpoint, or perspective, then the content is perceptually objective.

In summary, when the Boring Principle is imbedded within the information-processing framework, the content of an internal representation is whatever is invariant in the stimulus when tokening of the representation is invariant. If the invariance that is exhibited is with respect to viewpoint or perspective, then the content is perceptually objective. Now let us see how we can use this conclusion to show that the content is also perceptually robust.

4.3.2 Underdetermination, Variability, and Perceptual Robustness

If the content of an internal representation is perceptually objective, and the sensory input that activates the representation are both underdetermining and variable with respect to their distal causes, then this entails that wild tokenings of the representation are possible. And if wild tokenings are possible, then the content of the representation must be perceptually robust. My argument here is best illustrated by adapting a classic model from signal detection theory (Green and Swets, 1966; Figure 4.2).

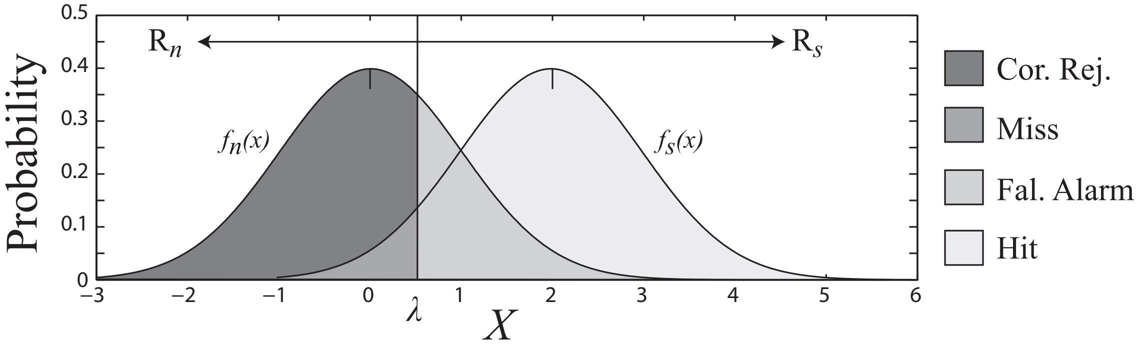


Figure 4.2: Adapted Model from Signal Detection Theory

To simplify matters, imagine an abstract characterization of how a simple sensory system enters into one of two possible internal states, R_n and R_s, depending on a proximal stimulation, x, caused by one of two distal states of the world: a "noise" state, N, and a "signal" state S. Suppose that each x is a value on a single dimension X, along which are positioned two probability density functions, $f_n(x)$ and $f_s(x)$, which reflect the probability of N and S causing some value x along the dimension. Also along the dimension is a threshold or boundary, λ, which determines the dependency between values along X and the internal states R_n and R_s. For example, in Figure 4.2, the mapping is such that when $x < \lambda$ the system enters into R_n and when $x > \lambda$ it enters into R_s.

Now suppose that R_s is a state within the representational scheme of our system, with S as its distal content, and that this content is (by hypothesis) perceptually objective, since tokening of R_s is invariant with respect to values along X. It follows that the content of R_s is also perceptually robust, under this simple picture. The reason is that the mapping between S and X exhibits both features of the mapping from the distal to the proximal that are central to the main narrative in vision science. First, the variation problem is captured in the fact that both distal states can cause a range of sensory inputs, as reflected by the probability distributions along X. Second, the underdetermination problem is captured in the fact that the two distributions overlap to a significant degree.[8] Thus, for given a value of X there is no exclusive mapping back to states of the distal world. This entails that, given any partitioning of X, wild tokenings of R_s will be possible—they will occur whenever N causes $x > \lambda$ (i.e., a "false alarm"). Likewise, there will be conditions when the signal will fail to cause R_s to be tokened, when S causes $x < \lambda$ (i.e., a

[8] Indeed, under the usual assumption of two Gaussian distributions, they both span the full length of X.

"miss"). Thus given perceptual objectivity, and assuming the underdetermination problem and variation problem hold, the content of R_s will be perceptually robust.

This picture is simple and idealized, but it reflects something basic about the relationship between perceptual objectivity and perceptual robustness in the visual system. Of course, sensory inputs are not uni-dimensional, but the same argument can be made using multidimensional extensions of signal detection theory (e.g., Ashby and Townsend, 1986). And it is not unusual to characterize both the underdetermining and variable nature of the mapping between the distal and proximal using probability distributions; indeed, this is the standard practice with probabilistic models of vision. Nor is it unusual to use a decision threshold along a dimension unusual as a means for characterizing the mapping from proximal stimulation to the tokening of an internal representation (Feldman, 2012; Godfrey-Smith, 1991).

So while this picture is idealized, the conclusion it leads to is quite concrete: if perceptual objectivity can be established, and the underdetermination problem and variation problem are assumed, then it follows that the content of the internal representation will be perceptually robust. Hence, if an objective reading for some constancy phenomenon in vision is warranted, then the internal representation posited to explain the constancy will have—indeed *must* have—all the crucial ingredients for being perceptual.

4.3.3 The Structure of the Argument

Putting the two strands together, we get the following *argument from constancy*, for some set of internal states of a representational scheme in the visual system (i.e., a set of internal representations), and some set of transformations of sensory inputs that cause tokenings of these representations.

> (I1) The contents of the internal representations are invariant with respect to the transformations.
>
> (I2) If the contents of the internal representations are invariant with respect to the transformations, then the contents are perceptually objective.
>
> (I3) Therefore, the contents of the internal representations are perceptually objective. [from (I1) and (I2)]
>
> (I4) The visual inputs that cause tokenings of the internal representations are underdetermining and variable.
>
> (I5) If the contents of the internal representations are perceptually objective, and the visual inputs that cause tokenings of the internal representations are underdetermining and variable, then the content is perceptually robust.
>
> (I6) Therefore, the contents of the internal representations is perceptually robust. [from (I3), (I4), and (I5)]

(I7) If the contents of the internal representations are perceptually objective and robust, then they are perceptual representations. [from (I6) and background assumptions]

(I8) Therefore, the internal representations are perceptual representations.

Let me unpack these premises. Premise (I1) of the argument is empirical: it simply references some constancy phenomenon, and the set of transformations with respect to which perception is invariant. Premise (I2) is the crucial premise, since it corresponds to the objective reading of the facts in (I1), and allows for a conclusion about perceptual objectivity (I3). Premise (I4) is just the (empirical) assumption that the underdetermination problem and variation problem obtain. Premise (I5) is the upshot of my argument illustrated using signal detection theory, which allows an inference to perceptual robustness (I6). (I7) rests on the assumption that the other staple and special ingredients are already satisfied by the representational scheme in question; thus, it follows that its internal states are perceptual representations (I8).

If an argument from constancy succeeds, then we have all that we need for defending the claim that perceptual representations are indispensable to explanations of perceptual constancy from within the information-processing framework—in line with Steps 1 - 4 of my argumentative strategy in Chapter 2.

First, we already had on hand the requisite recipe for perceptual representation. Second, we have some idea of explanatory role. Within the framework, internal representations are posited to explain perceptual constancies. Theories and models are also offered to explain how it is that these internal representations can exhibit invariance with respect to certain transformations of the visual input, given that the visual system must "overcome" the underdetermination and variation problems. These constancy-explaining internal representations have the sort of "dual" explanatory role that I mentioned in Chapter 2: they are posited to explain a phenomenon, but their structure and functional organization are also something to be explained. Third, in order to play this role, these posits must be an internal representation with content that is invariant to the relevant transformations, and which overcomes the underdetermination and variation problems. But fourth, as we have just seen, these requirements entail the key mental ingredients for my recipe for perceptual representation. Put differently, to posit internal representations to explain the invariance and specificity exhibited by a constancy phenomenon is to posit perceptual representations. For this reason, a notion of perceptual representation is indispensable.

All of the above trades, however, on whether an objective reading *is* warranted for the constancy phenomenon in question (and as we shall see in the next chapter, something must still be said to address the Job Description Challenge). One possibility is that *any* perceptual constancy exhibited by vision is sufficient for running the argument; that is, all forms of perceptual constancy in vision warrant an objective reading. Under this *broad version* of the argument, so long as invariance is exhibited with respect to some collection of transformations of visual input, it is distally specific—premise (I2) is true. An alternative possibility is that only some perceptual constancies are sufficient for running the argument. Perhaps an objective reading is warranted for some constancies in vision, and a subjective reading for others. Under this *narrow version* of the argument, only invariance with respect to some sorts of transformations are enough to establish distal specificity. Of course, a third possibility is that the argument never works, because an objective reading of perceptual constancies is never warranted.

So we have an explicit construction of the argument from constancy. Now let us look at why the argument might fail.

4.4 Cautionary Tales? Three Dependence Challenges for the Argument From Constancy

The argument from constancy holds some temptation for philosophers. For example, Dretske (1981, p.165 emphasis in original) appears to use the argument to show that representational content is distally specific:

> It is the *fact* of constancy. . . that accounts for the fact that our sensory experience gives primary representation to the properties of distal objects and not to the properties of those more proximal events on which it (causally) depends. It is this fact that explains why we see physical objects and not the effects that these objects have on our perceptual systems.

The temptation, exemplified by this quote from Dretske, is that the mere fact that there are perceptual constancies, and that internal representations are posited to explain them, warrants concluding that they must therefore be perceptual representations.

We should view such arguments with caution. In this section I introduce three prima facie challenges for any argument from constancy one might make. I refer to these as "Dependence Challenges", since they each raise questions about whether perceptual constancies in some way depend on facts that undermine their suitability for running an argument from constancy; that is, each gives us a reason for doubting that it is necessary to posit perceptual representations in order to explain constancy phenomena. While not decisive, I do think these challenges give us grounds for thinking that a narrow version of the argument, focused on object constancy, is the one most likely to succeed.

4.4.1 The Stimulus-Dependence Challenge

The first challenge is straightforward. The argument from constancy begins with (I1), the factual premise that some aspect of visual perception—the perception of object size, shape, color, or identity—remains invariant with respect to some set of transformations of the sensory input. If the premise is false, then the argument does not go through. And in general, it is not obvious that the factual premise is true. Indeed, the very possibility (and plausibility) of subjective readings suggests this is so. Consider two examples, color constancy and size constancy.

Perceived color depends on both spectral reflectance properties and illumination of surfaces. As generally understood, color constancy is supposed to obtain with respect to surface illumination: under transformations of the illumination of a surface or object, the perceived color of the surface remains invariant. However the available psychophysical data does not provide strong evidence that perceived (absolute) color is invariant with respect to transformations of illumination. For example, the main paradigms for investigating color constancy—color naming and surface color matching—do not provide clear evidence of color invariance with respect to changes in illumination (Foster, 2003). Color, one might think, is a special case, since it is debatable whether colors are even real properties of surfaces.[9] However the same moral applies when it comes to the facts about size constancy. A great deal of the early work on perceptual constancies was focused on size constancy with respect to changes in perceived distance. Indeed, it was this research that was interpreted by some as warranting an objective reading of size constancy (Boring, 1946; Brunswik, 1940; Thouless, 1931). In particular, the "Size-Distance Invariance Hypothesis", which states that the apparent size of an object is uniquely determined by the relationship between the visual angle of a stimulus and its apparent distance, was the subject of a considerable amount of research in the first half of the 20th Century. The classic version of this hypothesis has little to recommend it (Ross and Plug, 1998, p.521). For example, in an early review Epstein et al. (1961) already acknowledged that the hypothesis, which was motivated by principles of Euclidean geometry, is false.

[9] Arguments from constancy are popular amongst color realists in the philosophy of perception (Hilbert and Byrne, 2003; Tye, 2000). Cohen (2008) has argued against this approach in a manner similar to the Dependency Challenges I enumerate here.

The above does not show of course that there is not some degree of size constancy (or color constancy). After all, when Mr.Muscles walks towards me, his size appears constant, as does the color of his coat. The point is simply that claims about constancy in perception are empirical, and the intuitive fact that various types of constancy occur (to some degree) is not sufficient evidence for thinking the first premise of an argument from constancy is true. For if perception only exhibits a limited amount of invariance, then by the Boring Principle, a subjective reading might offer a better account of the available data. In which case, one need not posit perceptual representations to explain the phenomenon.

Visual illusions, which were relied on as evidence for a subjective reading of a perceptual constancy, also raise the "Stimulus-Dependence Challenge". For if perception remains invariant across both veridical and illusory perception, then this fact suggests that the object of perception is a higher-order property of the sensory signal, and not a property of something in the distal world. Indeed, at this point, one might even ask why illusions are not decisive in undermining the argument from constancy. The reason they are not decisive is that applying the Boring Principle to visual illusions does not obviously entail a subjective reading. Let me explain.

First, visual illusions are heavily reliant on constrained viewing conditions, or perspective. Many visual illusions, such as classic geometric illusions such as the Muller-Lyer Illusion or the Ebbinghaus illusion, as well as many color illusions, disappear under changes of viewing context. Consider the classic Ames room, which results in distorted perception of relative size of objects in a room.[10] The illusion

[10] These rooms are constructed so that the wall furthest from the viewer is unequal in size and

is not invariant with respect to transformations of perspective. As I will discuss at greater length in the next chapter, if one is allowed to navigate the room, or observe objects in the room interacting, the illusion fades (Kilpatrick, 1954). Under the Boring Principle, the object of perception is a function of the stimulus that is invariant when perception is invariant. A continuous transformation into or out of illusion inducing viewing conditions will result in a change in perception, which would suggest that there is in fact a change in the object of perception, under the principle. So even though perception in some "veridical" and illusory conditions might be the same, the application of the Boring Principle to illusory cases does not (at least straightforwardly) warrant a subjective reading.

Second, some apparent illusions are highly invariant, but are not obviously illusions. For example, consider the "bent-stick" effect, which occurs when we look at an object partially immersed in water. One might wonder whether in such a case we are genuinely having an illusory perception of object shape, or largely accurate perception of how an object looks during immersion. In other cases, the illusion is invariant, but we might consider it an anomalous case. Consider our perception of the Moon. On the one hand, the Moon Illusion, in which it appears larger near the horizon than when high in the sky (despite actually being *farther* from the viewer when it is at the horizon), is quite invariant with respect to viewing conditions. On

viewing distance, but in a manner to hide geometric cues of the room that might betray this fact. For observers in the appropriate position, the wall is perceived as equidistant at all points, and when objects are placed at different positions in front of the wall, the perception of relative size is distorted.

the other hand, the perceived size of the moon is, in general, closer to its retinal size, than its actual size (Holway and Boring, 1941). So whatever the perceived size of the Moon is, it is does not appear closely related to its actual size. However it is not implausible that celestial bodies are anomalous cases, which should not be considered representative of size constancy more generally.[11] So just as constancy achievement is not obviously decisive in supporting an objective reading, and hence an argument from constancy, neither is it obvious that facts about visual illusions are decisive in support of a subjective reading.

The upshot of the Stimulus-Dependence Challenge, then, is that the first premise of the argument from constancy cannot be taken for granted. Matters will depend on the empirical details about the extent and conditions of the invariance in question.

4.4.2 The Dimension-Dependence Challenge

Suppose we grant that the Stimulus-Dependence Challenge can be met, and that (I1) is true. With the empirical facts in, certain constancy phenomena exhibit complete invariance with respect to certain transformations of the visual stimulus: perceived color is wholly invariant with respect to some transformations of surface illumination; perceived size is wholly invariant with respect to some changes in viewing distance or retinal size. Even supposing this were true, we would still have grounds for doubting an objective reading. The reason is that many perceptual

[11] In this respect, the perception of celestial bodies might be "monster" anomalies, which can be quarantined from our theories regarding the objects of perception (Darden, 1991; Elliott, 2004).

constancies are relative to some *dimension* of the visual stimulus, and it is not obvious that invariance with respect to these dimensions suffices for making the object of perception non-perspectival. As discussed in Chapter 2, perspective is multi-dimensional: for vision a perspective, or viewpoint, involves conditions of illumination, viewing distance, retinal location, relative position or orientation, and the like. However, many constancies are relative to one of these dimensions. The Dimension-Dependence Challenge shows that we cannot simply plug in perceptual constancies into the argument from constancy, without consideration of what sort of constancy phenomena might actual exhibit invariance with respect to perspective—even granting premise (I1), we can still question premise (I2). The challenge gives strong reason to question whether a broad version of the argument from constancy is likely to succeed. For if the constancy is not with respect to visual perspective (which is multi-dimensional), it is not clear why we need to posit a perceptual representation to explain it.

One might counter that it is well-known that visual constancies are context-dependent. Consider again the examples of color and size constancy. If I illuminate a color patch in an otherwise dark room, without any other colored surfaces for comparison, color constancy will easily breakdown when I adjust the spectral properties of the illuminating light. Likewise, without cues to distance, perceived size will breakdown. Thus, while color constancy and size constancy pertain to a single dimension, our perception of these dimensions depends on our perception of the overall visual scene. However, this rejoinder is only on point if the context-dependence entails that the constancies are in fact invariant with respect to perspective, or viewpoint. In fact, pointing out the context-dependence of perceptual constancies points toward an even more serious challenge for the argument from constancy.

4.4.3 The Object-Dependence Challenge

In Chapter 2 I stated that perceptual content involves attributing a property to a particular, and constancy phenomena seem to arise with respect to the perceived properties of particulars: it is the object in front of me, my brother's cat, who I perceive as constant in size, shape, color, and identity, as he walks across the room. Described in this way, I *assume* that the particular is indeed something distal, an object in the world. But this is to in large part beg the question against a subjective reading. Constancy phenomena are always with respect to a perceived particular—there must be *something* that I am perceiving a certain way (Davies, 1991). Either the particular is itself a function of the sensory signal (a subjective reading), or something in the distal world (an objective reading). If the former, then an argument from constancy cannot succeed, for it is hard to see how we can have a subjective reading for particulars, but an objective reading of their constancies. If the latter, then we need an argument for why the objective reading for particulars is warranted. The Object-Dependence Challenge, then, is to provide an account of the particulars in vision consistent with an objective reading of perceptual constancies.

Meeting the challenge may seem tantamount to answering the question: what is a "visual object"? This is the sort of question vision scientists do not like to answer (or ask). For example, when broaching the subject Marr (1982, p.270) adopted the pessimistic conclusion that "anything" in the visual scene can qualify as an object. Such pessimism is not necessarily warranted, since there is sufficient psychophysical evidence to suggest that the visual system is so non-committal—as revealed, for example, by research on object-based attention, as discussed in the next chapter (Scholl, 2001). However, it is not guaranteed that a more systematic account will save the argument from constancy. For example, Feldman (2003) has argued that we should identify visual objects as a certain kind of "joint" in the tree-structure organization of complex visual features. If Feldman's proposal is on the right track, then visual objects are no more than a higher-order property of the sensory signal, and we end up with a subjective reading of perceptual constancies.

I think the Object-Dependence Challenge shows why a broad version of the argument from constancy is not going to work. One first needs to establish that the particulars to which our visual system attributes size, shape, color, and identity are indeed entities in the distal world. If such a fact is established, then one already has grounds for thinking that, at least to some extent, the relevant aspect of vision is perceptually objective. So we need an argument for why object vision is perceptually objective, and here a narrow argument from constancy might be useful, so long as it does not run afoul of either the Stimulus-Dependence or Dimension-Dependence Challenge.

4.4.4 Summary

The Dependence Challenges show that it is not enough to simply point to the fact that there are perceptual constancies, assume that objectivity follows, and that positing perceptual representations is required to explain them. First, the Stimulus-Dependence Challenge calls into question factual premise (I1), and whether some aspect of visual perception is indeed invariant with respect to certain transformations of the visual input. Second, even assuming premise (I1) is true, the Dimension-Dependence Challenge calls into question premise (I2), and whether the form of invariance on offer is with respect to perspective. Third and finally, the Object-Dependence Challenge raises a problem for the strategy in general, since all perceptual constancies depend in part on object vision, and thus matters rest on whether object vision itself demands a kind of objective reading.

What I think this all shows is that a broad version of the argument, for which any constancy will do, is unlikely to succeed. One strategy is to run an argument from object constancy, in order to address the Object-Dependence Challenge head on. Indeed, this is the approach I adopt in the next chapter. In contrast, Burge (2010) has recently defended a broad form of the argument from constancy. So let us see now see how his argument fairs with respect to the challenges I have laid out.

4.5 Burge On Objectivity in Perception

Recently, Burge (2010) has argued that perception affords the most primitive form of objectivity in the mind. Burge's argument is inspired by empirical research in perceptual psychology, for which he thinks a notion of perceptual representation is indeed indispensable. At its core, Burge's argument takes the form of a rather broad argument from constancy. Burge is not the first philosopher to make such an argument. For example, as mentioned earlier, Dretske (1981) points to an objective reading of perceptual constancies as the basis for his claim that the content of perceptual representations are entities in the distal world. However, Burge's approach is similar to (and indeed provides some of the inspiration for) my own. Thus it is worth assessing his argument in light of the dependency challenges I have enumerated and seeing, ultimately, why it fails.

In this section, I first summarize Burge's own recipe for perceptual representation. I then turn to the details of his argument from constancy. Before continuing, it is worth mentioning that Burge's views of intentional content, mental representation, and perception, are rather complex. Here I only discuss those details relevant to evaluating his argument from constancy.

4.5.1 An Alternative Recipe for Perceptual Representation

Burge is careful to articulate a conception of perceptual representation that he takes to be "non-deflationary".[12]. For Burge, the two crucial constitutive ingredients for perceptual representation are their objectivity, and representational function. It

[12] Burge thinks that theories of intentional content are deflationary because they entail some kind of reductionism that is incompatible with the explanatory practices of perceptual psychology. Such considerations are the motivation for his rejection of the explanatory relevance of theories of content to cognitive science (Burge, 2010, p.292)

is these ingredients that allow for a contrast between perceptual representation of the distal world and sensory registration of proximal stimuli.

When it comes to perceptual representations, Burge takes their content to be objective in the two senses I articulated in Chapter 2: they attribute non-perspectival, mind-independent properties to the distal world. He refers to such perceptual objectivity as being the "product of objectification":

> *Objectification* is formation of a state with a representational content that is *as of* a subject matter beyond idiosyncratic, proximal, or subjective features of the individual. The relevant subject matter is subject matter that is objective in one or both of the senses laid out... the subject matter is mind independent, or it is constitutively non-perspectival. Basically, the subject matter is comprised of entities in the physical environment. Objectification, then, is the formation of a representational state that represents the physical environment, beyond the individual's local, idiosyncratic, or subjective features. (p.396, emphasis in original)

[13]

According to Burge (p.309),[14] the representational function of the internal states of perceptual systems is to veridically represent the distal world. Thus by achieving objectification internal states of a perceptual system are able to veridically represent the distal world, given the limitations of the system (p.312). While

[13] The "as of" locution utilized by Burge is his preferred expression for conveying the idea that the relevant subject matter need not exist.

[14] Page numbers, when presented alone, are for (Burge, 2010).

representational function is thus tied to a notion of success, it is not to be associated with biological function. Rather the notion of representational function is supposed to be grounded in the explanatory needs of perceptual psychology, including vision science:

> The roles for both biological and representational functions are constitutively associated with success, and such functions ground explanations of success. Biological function grounds explanation of fitness, or successful survival for mating. Representational function grounds a distinctive sort of explanation: explanations of approximately veridical perception—and of failures of approximate veridicality—of the environment. (p.310)

Even though these notions of function are distinct, the representational functions of perceptual systems still operate in the service of the biological functions of organisms (Burge, 2003, p.510). Although veridical perception is not necessary for improving fitness, it can be sufficient.

These two constitutive ingredients help to mark the distinction between perceptual representation and sensory registration. For example, Burge thinks it is clear that the internal states of the sensory systems of simple organisms do not have these constitutive ingredients:

> Organisms like bacteria, amoebae, paramecia, worms, molluscs, clams are differentially sensitive to various attributes in the physical environment. They discriminate those attributes. Their sensory capacities carry information. They function to respond in certain ways, given this infor-

> mation. These organisms can discriminate light, heat, magnetic force, and so on. Responses to such discriminations function to enable them to live, move, and reproduce in their environmental niches. These sensory capacities are not perceptual. (p.315-316)

The internal states of these simple organisms do "functionally carry information" about their environment in the following sense:

> Some states that carry information that correlates with other states, and are causally dependent on them, have a function that capitalizes on such dependence. Broadly speaking, such states have such a function by virtue of having been selected through evolution, or perhaps designed as artifacts, partly because of the causal roles they play given the information that they carry. (p.317)

However, functionally carrying information is insufficient for perceptual representation, since it is insufficient for either objectification or representational function. First, the discriminative capacities do not suffice for objectification because:

> ...nothing in the individual's capacities...distinguishes (a) environmental causes that figure functionally in the individual's basic needs and activities from (b) sensory registration (or functional encoding) of proximal causes—from the surface effects of the environmental causes. (p.317)

Since these organisms cannot discriminate the proximal and the distal, they therefore cannot have objective, non-perspectival perception of the distal world.

Such sensory registrations amounts to mere "statistical, nomic, counterfactual, or causal relations" (p.400) between the sensory states and states of the environment, but does not involve attributing properties to particulars in the world.

Second, since there is no constitutive relationship between biological function and representational function—perceptual success need not be biological success—the sort of sensory discrimination of these organisms, which contributes to biological success, does not warrant a (non-vacuous) appeal to veridicality conditions necessary for representational function; that is, biological explanation of functional sensitivity to environmental conditions is not psychological explanation in terms of veridicality conditions.

To illustrate, consider the example of the Australian Jewel Beetle. The male jewel beetle can identify combinations of visual features that are indicative of the fitness of females of the species, but which make stubby beer bottles look immensely attractive. In the wild, the beetles will attempt to copulate with the bottles, since they appear as highly desirable mates (Gwynne and Rentz, 1983). So the visual system of the jewel beetle can detect certain feature combinations, and this serves a biological function, but it is incapable of discriminating between the property of being a female jewel beetle and being a brown dimpled object. In Burge's terms, the discriminative capacities of the jewel beetle are likely insufficient for objectification, and any appeal to veridicality conditions to explain its behavior would be vacuous. The visual system of the jewel beetle does not contain perceptual representations of mates per se.[15]

[15] This, at any rate, is a gloss one could give of the case in line with Burge's recipe. One might

In summary, given the ingredients for perceptual representation Burge identifies, they do not reduce to simply "carrying information, or any other sort of sensory discrimination, together with biological function" (p.316). In this way Burge draws a distinction between perceptual representation on the one hand, and sensory registration coupled with biological function on the other.

4.5.2 A Broad Argument from Constancy

In contrast to biological explanations of the sensory systems of simple organisms, Burge believes that perceptual psychology, and vision science in particular, make non-trivial appeals to perceptual representations (p.87-92). Burge bases his argument on themes in the main narrative in vision science, which he thinks demand an objective reading of perceptual constancies.

Burge identifies the main challenge in vision science as that of explaining how the visual system overcomes the underdetermination problem, and allows us to perceive the world:

> The primary problem for the psychology of visual perception is to explain how perceptual states that are of and as of the environment are formed from the immediate effects of proximal stimulation—principally from registration and spectral properties of the eyes.(p.89)

The visual system overcomes the underdetermination problem via processes

wonder whether, assuming the beetle is in fact attributing a property to a distal particular, why the content of its internal representation is not at least objective. It may well be, but then one needs to show the attribution is in fact to a distal particular.

of unconscious inference and what Burge terms "formation principles", which is his term for the idea of internalizing constraints (like Marr and Hildreth's spatial coincidence assumption):

> The transformations operate under certain principles that describe psychological laws or law-like patterns. These laws or law-like processes serve to privilege certain among the possible environmental causes over others. The net effect of the privileging is to make the predetermining proximal stimulation trigger a perceptual state that represents the distal cause to be, in most cases, exactly one of the many possible distal causes that are compatible with (but not determined by) the given proximal stimulation. I call psychological principles that describe, in an explanatory way, these laws or law-like patterns *formation principles.* (p.92, emphasis in original)

It is due to such formation principles that the visual system is able to distinguish the distal and the proximal, and achieve objectification (p.395). While acknowledging that vision science aims to explain both illusory and veridical perception (p.342), ultimately Burge thinks that appeal to perceptual representations are indispensable to the explanatory practices of the science; that is, perception is veridical enough of the time to warrant explanation in terms of representations with veridicality conditions:

> . . . [vision science assumes] that individual's perceptions are *approximately* accurate with respect to some environmental particulars and at-

> tributes *enough* of the time to ground a form of explanation that takes states with veridicality conditions to be the product and participants in the law-like formation patterns being explained. (p.88, emphasis in original)

So it should be clear that Burge wholeheartedly endorses much the same narrative about vision that I described earlier in this chapter. According to Burge, a distinction between sensory registration and perceptual representation is indispensable to the general explanatory practices of vision science. The reason is that unlike in the cases of the simple organisms described earlier, these practices make fruitful, and non-trivial appeal to perceptual representations in the explanation of visual phenomena. For Burge, the explanatory difference maker is perceptual constancies.

Burge describes perceptual constancies as follows:

> Perceptual constancies are capacities [sic] systematically to represent a particular or an attribute as the same despite significant variations in registration of proximal stimulation. . . these capacities cannot be explained simply as generalized weightings of registration of proximal stimulation They must involve principles for forming representation of specific environmental particulars and attributes. . . The intuitive idea of the constancies is that under different perspectives, a perceiver can represent a given particular or attribute as the same.(p.408)

Burge does not intend this as a definition, since it involves overt reference to representations, which (he thinks) would make the definition circular. Despite such overt reference, he still thinks that perceptual constancies provide the basis by which vision science makes non-trivial appeal to perceptual representations. For Burge, perceptual constancies are the "central instances of perceptual objectification" (p.397; also p.409). The reason is the same one we encountered before, which is that perceptual constancies show that perception is non-perspectival, and so perceptually objective:

> Perceptual constancies give empirical point to a distinction between perspective and subject matter. . . Representational content or mode of presentation is to be distinguished from subject matter. Any perceivable particular, property, relation, or kind can be perceptually represented in many ways, constituting different perceptual perspectives on the representatum. In all cases of perceptual constancies, this multiplicity of perspectives on a given subject matter emerges. . . A given perceptual representatum (kind, property, relation, or particular) is represented as that representatum, even as it is presented in different ways, from different perceptual perspectives. These differences in perspective and representational content, or perceptual mode of presentation, are caused by variations in sensory registration of proximal stimulation.(p.411)

Burge is clearly making an argument from constancy. Furthermore, he holds that the argument works for the main forms of perceptual constancy in vision, including color, size, shape, and object identity (p.409-410). Thus he offers a very broad argument from constancy in defense of objectivity in perception. Now, how does his view fair with respect to the Dependence Challenges?

4.6 Burge's Difficulties with Dependence

While I am very sympathetic to Burge's approach, I do not think his broad argument from constancy succeeds. In this section I first outline why his argument appears to run afoul of the Dependence Challenges. I then look at how he tries to account for visual illusions, via the notion "relevant representational alternatives". However, far from providing a means to address the Dependence Challenges, this notion either makes his account internally inconsistent, or leaves the object-dependence challenge unanswered.

4.6.1 Taking Perceptual Constancies for Granted

In developing his argument from constancy, Burge seems to take an objective reading for granted, and does not even acknowledge the presence of subjective readings in the literature (e.g., Epstein, 1977). Thus, it should come as little surprise that he does not appear to have resources for addressing the Dependence Challenges to arguments from constancy. Let us go through each of the challenges in turn.

Consider the Stimulus-Dependence Challenge first. Burge does offer some machinery for addressing visual illusions, and I will turn to a discussion of this machinery shortly. For now, it is worth noting that although Burge extensively cites the empirical literature in vision science on various constancy phenomena, his arguments are to a large extent divorced from the actual empirical details regarding these phenomena. In fairness, Burge (p.413) does appear to acknowledge that mere appeal to perceptual constancies is not enough to make his argument go through:

> Perceptual constancies are paradigmatic marks of objectification. I think that their presence in a sensory system is necessary and sufficient for the system's being a perceptual system. Their presence is certainly sufficient for perception and objectivity…I conjecture that they are also necessary. Since they are not characterized independently of the notions *representation* and *perception*, one cannot use the notion *perceptual constancy* as an independent 'criterion' to determine when one has a case of genuine perceptual representation and when one has a case of non-representational sensory registration. Empirical theory must draw the distinction and identify cases of perceptual constancy. (p.413)

So Burge seems to hold that to the extent perceptual constancies allow us to identify genuine perceptual representation, matters will depend on the empirical details. However he does not delve into these details. We have already seen that there are reasons to worry that perceptual constancies are stimulus-dependent, thus without further argument, Burge has at best pointed out that *if* an objective reading of perceptual constancies is warranted—premise (I2)—then we have an argument for the indispensability of perceptual representations. In fact, one criticism of Burge has been that the empirical facts are not obviously in his favor (Ganson et al., 2014; Olin, 2014). So it would seem Burge may be lacking a convincing defense of the factual first premise that is required to get an argument from constancy off the ground.[16]

Next, regarding the Dimension-Dependence Challenge, Burge appears to take for granted that any perceptual constancy suffices for establishing perception as being non-perspectival. But as we have seen, many forms of perceptual constancy pertain to particular dimensions of visual perception, and one must provide an argument for why constancy with respect to only one (or some) of these dimensions suffices for establishing that the content of the internal representations in question are non-perspectival. Burge does not provide such an argument.

Finally, regarding the object-dependence challenge, I believe Burge does not realize there is a problem. Burge does discuss object constancy at length (p.437-465), rejecting arguments to the effect that object representations are solely the domain of central cognition. I will not get into the details of his discussion for two reasons. First, there are reasons to worry that object constancy is stimulus-dependent, which is an issue I address head on in the next chapter, when making my own argument.[17] Second, Burge's account of object constancy depends on his approach to addressing

[16] Whether he has or not, in general, provided such a defense is not something I propose to discuss here, for two reasons. First, since he has not addressed the more pressing issue of the Object-Dependence Challenge, that is enough to already call his more general argumentative strategy into question. And second, for object constancy (which is the case that interests us), many of the same sorts of facts he references are discussed in the next chapter.

[17] Burge regularly cites the same literature on visual object recognition I discuss in the next chapter, but fails to acknowledge that one of the central debates in the field has been the degree to which the representational schemes that underlie object recognition are indeed viewpoint *dependent*.

worries about illusions. This approach, I will now argue, either makes his view inconsistent, or does not address the Object-Dependence Challenge.

4.6.2 The Dubious Relevance of Relevant Representational Alternatives

Burge is aware of the sort of threat that illusory perception seems to present for his broad argument from constancy. If we acknowledge that the same constancy mechanisms are at play in instances of illusory perception, then one can make the claim that whatever the object of perception is, it is present in illusory perception. As I argued earlier, to the extent illusions present a challenge to an objective reading, matters will again depend on the character and extent of the illusion effects in question. Still, it is worth looking at the details of Burge's proposal, to see if it offers any help with the Dependence Challenges more generally. I will argue that it does not.

To avoid issues that might arise from illusions, and other possible counterexamples, Burge proposes the following "Principle of Relevant Representational Alternatives" in his discussion of object vision:

> (RRA) For an individual to perceptually indicate and attribute an attribute (kind, property, relation), the individual, or something in the individual's psychology, must be capable of distinguishing instances of that attribute from *relevant representational alternatives*. That is, the individual or psychology must be able to distinguish instances of that

> attribute from instances of other attributes that the individual can discriminate and that also ground explanation (in fact, these biological explanations) of the individual's needs and activities in its normal environment.(p.466)

Under Burge's view: "given that (RRA) is met, having perceptual constancies with respect to an attribute suffices to have a capacity to represent the attribute perceptually" (p.466). Thus, so long as constancy is achieved across conditions such that our visual system distinguishes between RRAs, then the internal states of the system that exhibit these constancies will be perceptually objective. Since circumstances that give rise to visual illusions are purportedly not RRAs, they do not present problematic cases for his argument. The background motivation for (RRA), which I quote at length, can be found in Burge's discussion of perceptual constancies:

> The representational contents of perceptual states are partly determined by patterns of causal interrelations, usually in evolutionary history, with attributes in the environment. These causal relations ground explanations of individuals' basic biological functions, principally activities... The causal relations supplement the individual's discriminatory abilities to make perceptual content of the perceiver's states specific to attributes in the environment... Thus objectification in perception is partly beholden to environmental "context" for the nature of its representations and for what it can and does represent.

> To put the point differently: individuals' discriminatory abilities operate in a *restricted* context of environmental alternatives. . . It is enough that the individual have perceptual capacities that discriminate environmental attributes within ranges that have figured causally in the formation of the states and that are relevant to biological needs and activities.
>
> Thus perceiving an instance of a shape as that shape requires an objectifying capacity. It requires a capacity to discriminate one shape from another. It further requires discriminating shapes from other *relevant* elements in the environment—such as colors, edges, textures. Perceiving bodies as bodies requires discriminating them from events and properties, including the shapes of the bodies. These are *relevant alternatives* in perceptual explanations because they also figure in explanations of individuals' biologically basic functions in fulfilling needs and activities, and because they are relevant to the individuation and formation of perceptual states. The *perceiver* need not be able to discriminate bodies from illusions, proximal stimulation, sensations, abstract kinds, undetached entity parts. . . [T]hey do not figure in natural biological explanations of functional individual needs and activity. The perceiver's objectifying discriminatory abilities determine the nature of his perceptual abilities only within this larger environmental and ethological framework. (p.407, emphasis in original)

So Burge's view (in the details) seems to be that exhibiting perceptual constancy across a range of circumstances, where these circumstances are individuated in relation to certain *biological functions*, are what accounts for objectification in perception. I do not believe the addition of (RRA) saves Burge's broad argument from constancy. The problem is that Burge faces a dilemma: either the relevant discriminative capacities referenced in (RRA) pertain to the distal world, or they do not. Neither alternative saves Burge's argument.

First, suppose that they do. It follows that Burge's argument begs the question. I have presumed that Burge is providing an argument for why we should believe that the objective reading for perceptual constancies is warranted. If (RRA) pertains to attributes of particulars in the distal environment, then it simply amounts to a statement of the desired conclusion: that the system in question perceptually represents the distal world. More specifically, under this first interpretation of (RRA), Burge begs the question with respect to the Object-Dependence Challenge; he presumes that the visual system can discriminate between particulars in the distal world.

Second, suppose that they do not. Then the discriminative capacities are simply those of sensory registration. Burge's argument no longer begs the question, but then it no longer leads to the desired conclusion. It is hard to square such an interpretation with his explicit rejection of the idea that sensory discrimination along with biological function suffices for perceptual representation. Indeed, if (RRA) only pertains to sensory discrimination, then he has dissolved the distinction between sensory registration and perceptual representation. No doubt simple organisms, such as the jewel beetle, exhibit some level of perceptual constancy with respect to their own environmental contexts. Thus, these organisms would seem to meet Burge's sufficient conditions for perceptual representation after all. This is not an outcome that I think Burge would welcome. Far from vindicating the explanatory practices of vision science, and perceptual psychology more broadly, this second interpretation of (RRA) risks rendering vacuous their appeal to perceptual representations.

So in short, I do not see any obvious way of interpreting (RRA) that does not wholly undermine Burge's argument from constancy. Thus it provides little help in addressing the Dependence Challenges. Burge aims to explicate why appeals to perceptual representations are indispensable to the explanatory practices of mainstream vision science. In order to realize this aim, Burge makes a very broad argument from constancy. I do not believe the argument succeeds.

4.7 Conclusion

Historically, perceptual constancies took center stage within debates regarding the objectivity of perception. I looked at how one might argue that perceptual representations are indispensable to the explanation of these constancies. Having reconstructed what I have referred to as the "argument from constancy", I then enumerated some Dependence Challenges for the argument. I went on to argue that Burge's (2010) version of the argument from constancy cannot overcome these challenges.

A reoccurring theme throughout my discussion has been that a broad version of the argument is unlikely to succeed—or at least, not without first determining whether the argument will succeed when it comes to object constancy. This was the upshot of the Object-Dependence Challenge. Thus, in the next chapter, based on research on visual object recognition (and object vision more generally), I develop an argument from *object* constancy that I believe overcomes the Dependence Challenges.

Chapter 5: Representing and Recognizing Objects

5.1 Introduction

In this chapter I argue that a notion of perceptual representation is indispensable to those practices in vision science that are aimed at explaining facts about visual object recognition. In other words, I provide an argument for why the Distal Object Thesis is indispensable to explanations of object recognition. My argument is an argument from constancy, which relies on facts about *object* constancy—namely, the viewpoint invariance of object recognition—to show that visual object representations are both perceptually objective and robust. As we saw in the last chapter, there are a number of Dependence Challenges that this sort of argument faces. In making my argument I believe I can meet these challenges.

The rest of the chapter is largely structured around the Dependence Challenges I have enumerated. In Subsection 2, I illustrate the central importance of viewpoint invariance and object specificity to research on visual object recognition, and lay out my argument from *object* constancy. I also spell out why, in virtue of resting on facts about viewpoint invariance, my argument is able to avoid the Dimension-Dependence Challenge. In Subsection 3, I look at how we should conceive of the "object" of object recognition. By incorporating research on object persistence and object-based attention, I show that my argument also meets the Object-Dependence Challenge. In Subsection 4, I discuss the substantial body of evidence showing that the representational schemes that underlie object recognition are in fact viewpoint *dependent*. I show that this viewpoint dependence is compatible with my argument, given the role of perceptual learning in recognition. Thus my argument also meets the Stimulus-Dependence Challenge. In Subsection 5, I discuss some of the less central ingredients for perceptual representation I listed in Chapter 2. Finally, In Subsection 6, I conclude the chapter.

5.2 Recognizing a Narrow Argument from Constancy

Like perceptual constancies, object recognition is one of the ever present facets of visual perception. The world we see is a world of objects, and generally speaking, we always see these objects *as* something. When I look at Mr.Muscles, I do not simply see a thing: I *categorize* him as a kind of thing (a cat), and I also *identify* him as a particular individual (my brother's cat). Most research on object recognition in vision science aims to explain how it is that we are able to both categorize and identify objects visually (Riesenhuber and Poggio, 2000). Marr (1982) believed that the representation and recognition of objects (or rather, their shape) was the chief function of the visual system. Attaching such functional significance to object recognition remains common today.[1] For example, Peissig and Tarr (2007, p.76) open their review of the previous 20 years of work in the field as follows:

[1] At least, amongst those who do research on the topic. For a critique of this consensus, see Cox (2014)

> At a functional level, visual object recognition is at the center of understanding how we think about what we see. [It] is a primary end state of visual processing and a critical precursor to interacting with and reasoning about the world.

The functional significance of object recognition derives from the fact that it affords *objective* perception of the world (Boring, 1946, p.107). It is at this sort of end stage of visual processing that the Distal Object Thesis finds a place: vision goes beyond the proximal stimulation, and reaches into the distal world. I believe that this connection between perceptual objectivity and object recognition is well-founded, and can be used to ground an argument from object constancy.

In this section, I first provide some evidence that viewpoint invariance and object specificity are considered the main explananda for research on object recognition. Since viewpoint invariance is invariance with respect to perspective, I use these facts about object recognition to develop an argument from constancy that readily avoids the Dimension-Dependence Challenge.

5.2.1 Viewpoint Invariance and Object Specificity

The central explananda for explanations of object recognition are of the typical sort for perceptual constancies: explaining some level of invariance and specificity in visual perception. In the case of research on object recognition, the central question is how recognition can be object specific to things in the distal world, while also being invariant with respect to transformations of viewpoint: changes in retinal size, position, illumination, orientation (rotation in both the picture-plane and depth-plane), and perceived distance. To explain these facts, internal representations for object identity and category membership—what I will call more simply *object representations*—are posited, and theories and models are developed to explain how these object representations can exhibit both viewpoint invariance, and object specificity. And it is the positing of this distal specificity that amounts to a commitment to the Distal Object Thesis.

That these are the central explananda in the field, and that object representations are posited to explain them, has remained fairly consistent over time. For example, much the same explananda for object recognition were acknowledged by Marr and Nishihara (1978, p.270-272), who pioneered research on object recognition. For Marr and Nishihara, any adequate explanation of recognition must explain how a representational scheme (in the sense from Chapter 3) for object shape can be constructed using the information available in the retinal image. Crucially, among other features, the content of the internal representations in the scheme must have appropriate uniqueness to objects of different shape (i.e., object specificity), but also stability given changes in orientation (i.e., invariance). Both explananda were also identified by Biederman (1987, p.117), in another landmark paper, as among the basic phenomena of object recognition.

More recently, researchers in the field have come to explicitly acknowledge that viewpoint invariance and object specificity provide the two most fundamental facts about object recognition that cry out for explanation in terms of internal representation. For it is not enough that object recognition exhibit either of these features. Rather it is their conjunction which is in demand of explanation, given that the variability problem is at its worst when it comes to changes in viewpoint:

> The main computational difficulty is the problem of variability. A vision system needs to generalize across huge variations in the appearance of an object such as a face, due for instance to viewpoint, illumination, or occlusion. At the same time, the system needs to maintain specificity. (Riesenhuber and Poggio, 2000, p.1199)

> The crux of the object recognition problem lies in the ability to produce a representation that can selectively identify individual objects in a manner that is essentially tolerant ("invariant") to changes in position, size, and context... From a computational perspective, constructing a representation that is either highly selective or highly tolerant is trivial; the challenge is to build a system that can produce a representation that is simultaneously selective and tolerant. (Rust and DiCarlo, 2010, p.12978)

While another branch of research on object recognition also focuses on the underdetermination problem for object vision (e.g., Kersten and Yuille, 2003), understanding how the brain constructs object representations given the high variability in input is perhaps the central question in the field—a field that includes the psychophysics and cognitive neuroscience of object recognition in humans and primates, as well as attempts to build computational models that mimic human performance, and which are constrained by neural architecture. The reason for this broad interest is that overcoming variability with respect to viewpoint makes object recognition an exceptionally computationally demanding task, as clearly stated by DiCarlo and Cox (2007, p.333):

> Object recognition is computationally difficult for many reasons, but the most fundamental is that any individual object can produce an infinite set of different images on the retina, due to variation in object position, scale, pose and illumination, and the presence of visual clutter.. . . Indeed, although we typically see an object many times, we effectively never see the same exact image on our retina twice. Although several computational efforts have attacked this so-called 'invariance problem',. . . a robust, real-world machine solution still evades us and we lack a satisfying understanding of how the problem is solved by the brain.

DiCarlo and Cox do not overstate the difficulty of "solving" object recognition. To illustrate, consider the performance of some common and state-of-the-art computational models of object recognition.

HMAX (for "hierarchical model and X") is a commonly used brain-inspired computational model of the stages of visual processing (Riesenhuber and Poggio, 1999; Serre et al., 2007). The end stage of HMAX is a neural network layer designed to mimic aspects of the functional organization of inferior temporal cortex (IT), the brain region most commonly associated with object recognition in humans and other primates (DiCarlo et al., 2012). While widely used as a proxy model of object representations in human IT, the model easily deviates from human performance. Recently, Stojanoski and Cusack (2014) introduced a method for image scrambling that preserves the low-level properties of an image (such as spatial frequency), while rendering the object in the image unrecognizable (e.g. the scrambled beaver is not recognizable, when introduced in isolation). Crucially, Stojanoski and Cusack found that the "object representation" layer of the HMAX model showed the same activity to both the intact and scrambled image, despite the fact that psychophysical evidence showed a clear difference between the recognizability of the two images. So while it is inspired by the neurophysiology of the human brain, HMAX largely deviates from human performance.

More recently so-called "deep neural networks" have been recruited to investigate object recognition in the brain. In one recent study, Cadieu et al. (2014) compared an implementation of recent deep neural network models (Krizhevsky et al., 2012) to the organization of human IT. When trained to classify object exemplars that included variation of viewpoint, these models performed similarly to humans, and developed an internal organization that closely matched some of the large-scale organization of human IT. The results of Cadieu et al. are impressive, in that these state-of-the-art computational models both achieve human-like levels of object recognition performance, and also contain an organization that is highly similar to human IT. As impressive as these results are, deep neural networks have puzzling properties, and it is still unclear *what* it is they learn when they are trained to categorize natural images. For example, using genetic algorithms images can be evolved that would be characterized by deep neural networks as objects of particular categories, with a confidence level $> 99\%$ even though they are visually unintelligible to human viewers (Nguyen et al., 2014).[2]

These two computational approaches are not the toys of computer scientists who, ignorant of vision science, have attempted to solve a real-world problem of which they have little understanding. Both reflect years of research by experts in the field who have a detailed understanding of both the psychology and neuroscience of object recognition, as well as state-of-the-art methods from computer science. This understanding is built into the guts of these brain-inspired models. I believe the modest success of these models illustrates the impressive challenge of explaining the core facts of object recognition. By way of comparison, consider that the human brain is able to recognize objects as quickly as < 150 ms after stimulus onset (Kirchner and Thorpe, 2006; Thorpe et al., 1996). The speed of such“ultra-rapid” recognition is only slightly longer than the time it takes visual signals to travel along the neuronal pathways from the retina to IT cortex.

Research on visual object recognition aims to explain how the brain constructs representations of objects that are both viewpoint invariant, and object specific, given the extreme computational challenge of the variation problem. This much, I believe, is widely accepted in the field. What is less commonly acknowledged is how these features of object recognition relate to the main narrative in the story of vision.[3] It is clear that researchers have in mind an objective reading when it

[2] For a video discussing how these stimuli were generated go to: https://www.youtube.com/watch?v=M2IebCN9Ht4.

[3] Indeed, I know of no one in the literature—with perhaps the exception of Burge (2010)—who has explicitly connected the very contemporary interest in the visual object recognition literature, with the very traditional issues regarding objectivity in perception in the last chapter.

comes to the object specificity of recognition, but what has not been noticed is that, through an argument from constancy, viewpoint invariance provides the *basis* for thinking that object representations are distally specific. I will now elaborate on this point.

5.2.2 The Argument from Object Constancy

viewpoint invariance of object recognition is the somewhat technical description for the fact of object constancy.[4] This kind of constancy, with respect to changes in viewing conditions and hence states of the retinal image, is the closest vision gets to being wholly non-perspectival. Thus, if object recognition is indeed viewpoint invariant, then we have a strong foundation for an argument from constancy.

In the last chapter we saw that if an objective reading of a perceptual constancy is warranted, then the content of the representations that are posited to explain the constancy will be perceptually objective, but then also perceptually robust. The reasoning to the first conclusion followed straightforwardly from the truth of an objective reading of a constancy. The second conclusion followed from the fact that, if (1) the content of some set of representations in vision is perceptual objectivity, and (2) the underdetermination and variability problems hold, then (3) wild tokenings will be possible, so the content must also be perceptually robust. This kind of argument looks promising, when we start with facts about object recognition.

Given their viewpoint invariance, the contents of object representations in the

[4] For example, in Walsh and Kulikowski (1998), a volume on perceptual constancy, the chapters on object constancy are primarily discussions of research on object recognition.

visual system are non-perspectival, and hence perceptually objective. Since this invariance results from overcoming both the variability and underdetermination problems, it also follows that the content is perceptually robust. Thus, assuming that they satisfy the other staple and special ingredients for perceptual representations, the content of the object representations that are posited to explain the central facts of object recognition have all the ingredients for genuine perceptual representations. Breaking the argument down explicitly, we get the following argument from *object* constancy, for the set of internal states of the representational scheme for objects in the visual system (i.e. object representations), and the transformations of viewpoint over which which their content is invariant:

> (O1) The contents of object representations are invariant with respect to transformations of viewpoint.
>
> (O2) If the contents of object representations are invariant with respect to transformations of viewpoint, then the contents are perceptually objective.
>
> (O3) Therefore, the contents of object representations are perceptually objective.
>
> (O4) The visual inputs that cause tokenings of object representations are underdetermining and variable.
>
> (O5) If the contents of object representations are perceptually objective, and the visual inputs that cause tokenings of the object representations

are underdetermining and variable, then the contents are perceptually robust.

(O6) Therefore, the contents of object representations are perceptually robust.

(O7) If the contents of object representations are perceptually objective and perceptually robust, then the representations are perceptual representations.

(O8) Therefore, object representations are perceptual representations.

This argument already avoids one of the Dependence Challenges from the previous chapter. Viewpoint invariance is invariance with respect to visual perspective: we recognize objects across transformations of orientation (such as rotation in plane and depth), spectral illumination, and retinal size and position (and hence position in the visual field and viewing distance). Thus the Dimension-Dependence Challenge is not a difficulty for my argument, since viewpoint invariance is invariance with respect to *all* dimensions of visual perspective—in other words, premise (O2) is true. However, arguably the Dimension-Dependence Challenge is the least daunting of the difficulties I have introduced, for it simply requires that an argument from constancy rest on invariance with respect to perspective (a reasonable demand, given how I have defined perceptual objectivity). So while a promising start, more work needs to be done to show that my argument from object constancy also avoids the other two dependence challenges.

5.3 Object Recognition and Object-Dependence

Part of my motivation for looking to object recognition as the foundation for an argument from constancy was to meet the Object-Dependence Challenge head on. The issue, recall, is that perceptual constancies typically pertain to attributes of some particular, which raises the question as to the ontological status of these entities. Are they a complex of visual features, or an entity in the distal world? If the former, then objective readings of perceptual constancies in vision may never be warranted, and perceptual representations need not be posited to explain them. If the latter, then we need an argument for why such an objective reading is appropriate. I suggested in the last chapter that an argument from constancy pertaining to object constancy might avoid the Object-Dependence Challenge. I have since proposed just such an argument. Now we need to see if it indeed meets the challenge.

In this section I begin by outlining why the organization of the recognition process (as it is typically characterized) does not straightforwardly provide a means of avoiding the Object-Dependence Challenge—indeed it seems to put the problem in starker relief. However as I go on to argue, once we consider explanations of object vision more generally, and appeal to work on object persistence and object-based attention, my argument from object constancy can indeed meet the challenge.

5.3.1 What is the “Object” of Recognition?

When vision scientists explain how we identify or categorize something in a visual scene, there is still the issue of what kind of “object” they presume to be speaking of: a combination of visual features, or a particular in the world? For all I have said so far, we have yet to see any reason to believe my argument can overcome the Object-Dependence Challenge. And when we consider how object recognition is characterized and explained as an information-processing task, it is clear that the challenge persists.

As I have already mentioned, the legacy of Marr looms large in the field of object recognition. The history of the field traces to (Marr and Nishihara, 1978, p.269), who applied Marr's information-processing approach to: "the problem of representing three-dimensional shapes for the purpose of recognition". While their proposal was aimed specifically at recognizing objects based on 3D shape, their approach was the catalyst for all subsequent research on object recognition. Following Marr and Nishihara, the process of recognition is typically characterized as a matching of some sort of internal representation that has been constructed online with some stored object representation of an individual identity or category (Figure 5.1; compare Liu et al., 1995, Fig.1). While differing in the details, both of the two main theoretical approaches in the field explain recognition as resulting from this sort of matching operation.

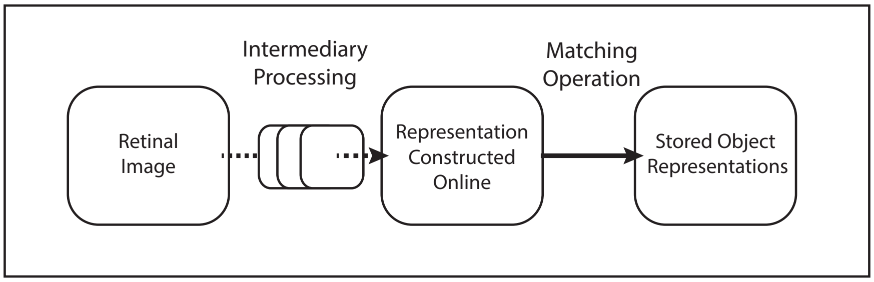

Intermediary Processing
Matching Operation
Retinal Image
Representation Constructed Online
Stored Object Representations

Figure 5.1: Object Recognition as an Information-Processing Task.

First, *object-centered* approaches originated with Marr and Nishihara, but have been most fully developed by Biederman (1987) in his “recognition-by-components” (RBC) theory (see also Hoffman and Richards (1984); Stankiewicz (2002)). What all object-centered approaches have in common is the idea that the representational schemes for recognition utilize an object-centered frame of reference that stores representations of the overall shape of objects. When we recognize an object from some viewpoint, a representation of the shape of the object is constructed from a set of volumetric primitives, which is then matched to the stored shape representations for different objects. Object-centered approaches can be understood as applying primarily, or even exclusively, to categories that can be recognized by shape alone, replying on structure relations between image features, and do not include a role for surface features, such as color or texture, in recognition.

Second, *view-centered* approaches largely developed as a reaction against the object-centered approaches of the 1980s (Bülthoff and Edelman, 1992; DiCarlo et al., 2012; Riesenhuber and Poggio, 1999; Tarr, 1995). What all view-centered approaches have in common is the idea that the representational scheme for recognition consists of stored representations that are in some sense “view-based”: a view-specific representation of an object is matched to stored representations of images for different individual object categories using (for example) an interpolation process. View-centered approaches readily incorporate all information in an image, including color and texture, but typically at the expense of ignoring information about the structural relationship between image features (a point that often goes unappreciated; Barenholtz and Tarr, 2006). Sometimes they also focus on only on stored “fragments" of an image (Ullman, 2007).

At present, view-centered approaches are something of the received view in research on visual object recognition (for discussion, see: Hayward, 2003, 2012; Stankiewicz, 2003). Whether this reception is warranted, what is important for present purposes is that both approaches, at their core, offer the same sort of explanation of the recognition process: that of matching a representation of an object that is constructed online, with those for identity and category that are housed in a long-term memory store in the visual system.[5]

For both kinds of approach, we can ask *what* they assume to be the particular that is being identified or categorized. Furthermore, both theoretical approaches appear compatible with either an objective or subjective reading. If the particular is identified with the content of the representation that is constructed online and

[5] As some have noted, it is not even entirely obvious that object-centered and view-centered approaches are even in direct competition (Hayward, 2003; Peissig and Tarr, 2007). For example, object-centered approaches were developed to explain recognition of objects based on 3D shape absent information from surface color or texture. In contrast, view-centered approaches have at least aspired to provide a general account of recognition, where all information from the retinal image is utilized—though often they tend to ignore or downplay the importance of structural relations between features of an image (Barenholtz and Tarr, 2006). Thus it is possible that rather than providing competing theories of object recognition, they reflect different ways in which (perhaps) diverse object representations are recruited given different task demands (Schyns, 1998). How I have differentiated between object-centered and view-centered views might also be an inaccurate way of characterizing the theoretical space. For example, Hayward (2012) considers Biederman's RBC theory to be a view-centered theory. A similar opinion is expressed by Stankiewicz (2003), who is sympathetic to RBC theory.

matched, then either it is itself distal, or it is not. If the latter, then it is unclear how we get viewpoint invariance, when the object is non-distal. If the former, then we again need an argument.

In summary, although the object representations that are posited to explain recognition are seemingly perceptual, and so involve attributing a property to a distal particular, the explanations offered of the recognition process do not seem to posit this sort of attribution. So if my argument from object constancy is to succeed, I need to a direct answer to the question: what is the object (i.e., particular) of object vision?

5.3.2 The Persistent Need for Object Indexes

Unfortunately the field of object recognition is not primarily concerned with the particulars of object vision. Fortunately, other areas of research on object vision are. And once we get clearer on the different aspects of object vision, I think the Object-Dependence Challenge can be met.

So far I have highlighted recognition as one aspect of object vision. A second aspect is what I will call object *discrimination*: the grouping together of features in the visual image into a single unit. Discrimination involves the processes of figure-ground segmentation, and the representation of 2D shape. Representations of visual objects in this sense are clearly perspectival. Feldman's (2003) proposal that we think of visual objects as a node in a visual feature hierarchy is a way to think of the "objects" of discrimination. A third aspect is what I will term object *individuation*: the visual perception of an individual, or particular, object in the visual scene. When we recognize (i.e., identify or categorize) objects we attribute a property to a particular individual. With respect to my argument from object constancy, the Object-Dependence Challenge amounts to the following question: is the particular of object individuation considered to be simply that of discrimination, or is it hypothesized to be something non-perspectival?

Most research on object individuation has been pursued under the labels "object persistence" and "object-based attention" (Scholl, 2001). This research is largely carried out independently of work on object recognition, even though the latter presupposes a notion of object "tokens" that can be "typed" as having a particular identity, or membership in some category (Treisman and Kanwisher, 1998). I believe that how researchers try to explain object persistence provides a means of meeting the Object-Dependence Challenge.

On the one hand, a good deal of evidence suggests that rather than being directed toward visual features, or regions of the visual field, much of visual attention is object-directed (Scholl, 2001). On the other, being able to attend to an object over time requires an ability to *track* it over time, despite changes in both spatial position, and visual features. Here is how Kahneman et al. (1992, p.177) describe the phenomenon:

> Imagine watching a strange man approaching down the street. As he reaches you and stops to greet you he suddenly becomes recognizable as a familiar friend who you had not expected to meet in this context.

> Throughout the episode, there was no doubt that a single individual was present; he preserved his unity (in the sense that he remained the *same* individual), although neither his retinal size, his shape, or his mental label remained constant."

Here our ability to track the stranger, as a single entity in our environment—to represent object *persistence*—reflects the fact that the brain somehow "solves" the *correspondence problem* (Ullman, 1979). A main focus of research on object persistence is explaining how the visual system overcomes the this problem, as noted by Flombaum et al. (2009, p.136):

> our visual system must constantly decide whether current stimulation reflects a novel object, or whether it corresponds to an object that was already encountered a moment ago. . . If we couldn't solve this type of problem, visual experience would be incoherent. This is a problem in part due to the sheer information load involved, and in part because a previously experienced object may be reencountered—even a moment later—in a different location. . . and/or with different visual surface features. How are these challenges overcome? How do we readily see objects as the *same* enduring individuals from moment to moment?

All explanations of how the brain overcomes this problem involve positing some form of *object indexes*, which have a few distinguishing features (Leslie et al., 1998, p.11):

I. They are token internal representations that function as "pointers" to objects.

II. They do not inherently represent any property or feature of the objects.

III. Indexing an object is a process for selectively focusing attention, and is therefore resource-limited.

IV. Indexes are assigned primarily based on perceived location in a visual scene.

While varying in the details, all theories of object indexes posit token internal representations that minimally have properties I-VI. For example, this is true of the most influential theories of object indexes, which posit "FINSTs" indexes (Pylyshyn, 1989), or "object file" indexes (Kahneman et al., 1992).[6] (I) and (II) are inspired by the idea of pointers in computer programming languages, which indicate a position in a data structure where some variable is stored. A crucial implication of (II) is that object indexes are something of a "blank slate". They are an abstract internal representation of a particular, and so are not intrinsically tied to any visual features, or spatial or temporal properties of an object. Rather, when an index is recruited, these properties are to be bound up with the representation. (III) reflects the fact that object indexes play a crucial role in object-based attention, the idea being that

[6] For present purposes, the details of these theories are not important. "FINST" is for "Fingers of INSTantiation", which is so-named because physically pointing at something with your finger is also somewhat analogous to the idea of an object index (Pylyshyn, 1989).

we can only keep a limited number of indexes online at once. (IV) specifies that indexes are applied to objects in locations, but not to locations themselves.

Leslie et al. (1998) also suggest some principles for index assignment: distinct objects can only have one index; once assigned, an index tracks an object through space; distinct indexes can be assigned to distinct objects iff they occupy distinct locations in space; and since they are a finite-resource, indexes must be reusable. For example, if attention must track a novel object, and all indexes are in use, then an old index must be recycled and applied to the new object.

Object indexes are the central posit in explanations of object persistence. If some change in the visual input is perceived (either because of external changes in the environment, or eye moment) the visual objects that are discriminated in the scene must be matched to already recruited object indexes, or new ones must be put into service. In order to maintain a stable representation of attended-to objects in the visual scene, and to solve the correspondence problem, the visual system must continuously carry out the following sorts of operations (Kahneman et al., 1992, p.179):

> V. A correspondence operation that determines whether a discriminated object is novel, or the target of an index that has moved to a new location.
>
> VI. A review operation that retrieves information about the target of an index based on its prior manifestation in the visual field.
>
> VII. A completion operation that uses the current representation and

reviewed information to produce a perception of change or motion linking the prior and current representations of the object.

Of most interest for our purposes is the the correspondence operation, and the information it relies on. Object indexes are abstract pointers to objects, and so are not intrinsically tied to information about an object. Nonetheless carrying out the operation requires matching stored information bound up with the index to the present visual input. The two available sources of information from the visual scene are spatiotemporal information, and featural information. As a crude comparison, these two sources of information amount to tracking based on where something is versus what it looks like. The dominant view in research on object persistence is that spatiotemporal information is the primary, if not exclusive, source of information for updating indexes (Flombaum et al., 2009; Mitroff and Alvarez, 2007). At the same time there is a good deal of research challenging this received view, suggesting that in many cases surface features can play an important, or dominant role in index updating (see e.g., Hein and Moore, 2012). This is very much a topic of ongoing research, and I of course will not take a stand one way or the other. Minimally, it is plausible that both sources of information are considered to play a role in achieving correspondence (Hollingworth and Franconeri, 2009; Moore et al., 2010), although spatiotemporal information may be more dominant (Leslie et al., 1998).

So far I have described the architecture that is posited to explain object persistence in vision, but without saying anything about how it is investigated. So before relating the architecture to explanations of recognition and the Object-Dependence

Challenge, I would like to give a sense of the sort of paradigms that have been used as the basis for positing, and investigating, object indexes. Here I will briefly describe two of them (Figure 5.2).[7]

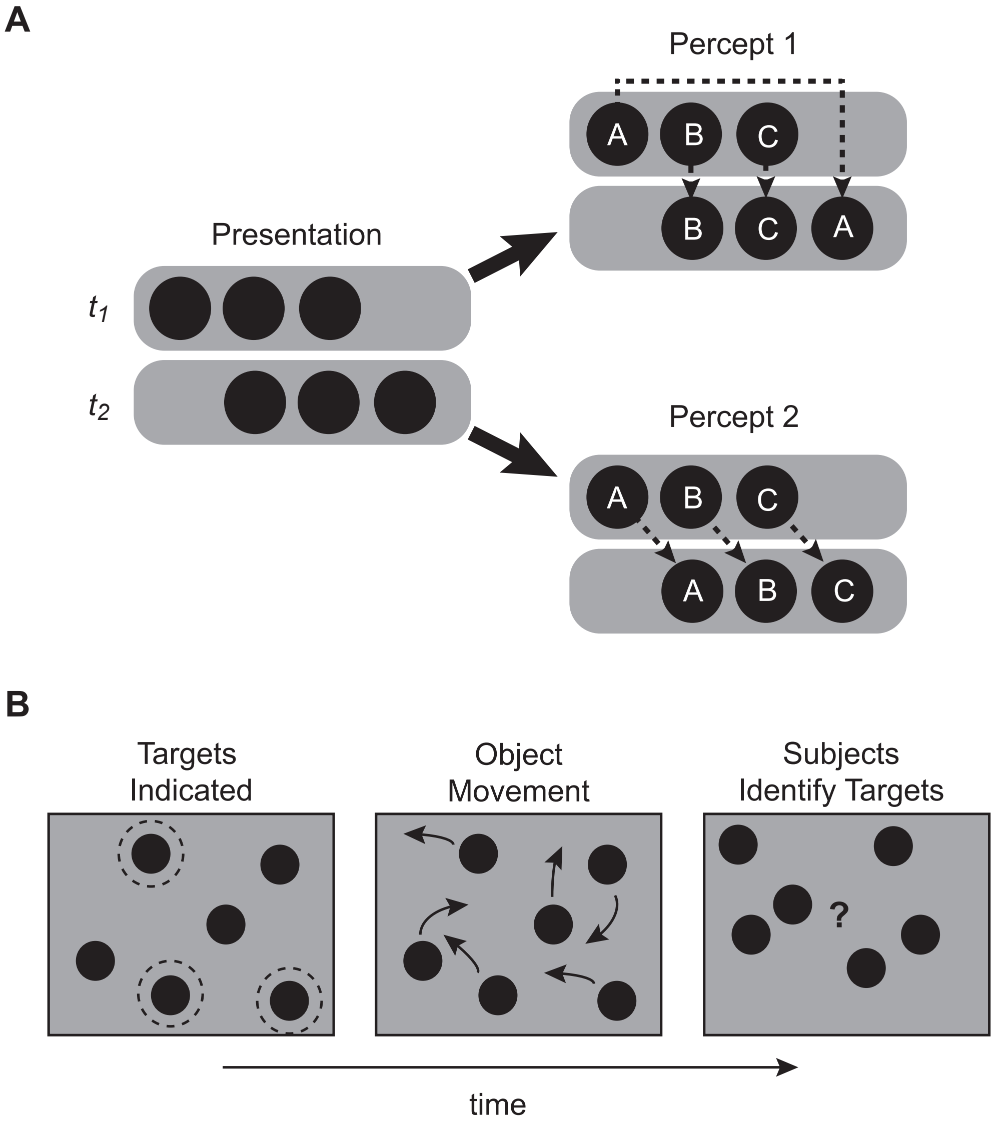


Figure 5.2: The Ternus Display and Multi-Object Tracking.

First, in a typical *Ternus Display* (Ternus, 1926) three black dots are presented

[7] Figure 5.2A is after Hein and Moore (2012, Fig 2), and Figure 5.2B is after Scholl (2009, Fig 2.1).

on a screen at time t_1, and then three black dots are presented at t_2 shifted to the side (Figure 5.2A). Depending on the length of delay between presentation times t_1 and t_2 subjects either perceive the left most element, A, as shifting to the right side of the string (Percept 1, shorter delays), or all three elements as shifting to the right while maintaining their order (Percept 2, longer delays).[8] Second, in the *multi-object tracking* paradigm (Figure 5.2B) multiple identical items are presented on the screen, and subjects are instructed to track a subset of them. The items then move across the screen, and after they have stopped, subjects must indicate which were the items they were supposed to attend to (Pylyshyn and Storm, 1988). While there are many variations of these paradigms to be found in the literature (see e.g., Hein and Moore, 2012; Scholl et al., 2001), each seems to involve object indexing. In the Ternus Display differences in the temporal properties of stimulus result in two different outcomes to the correspondence operation. In multi-object tracking indexes appear to be applied concurrently to multiple items which can only be differentiated by spatiotemporal information[9]

Now with a bit of a sense of the sorts of methods and phenomena used to investigate object persistence, let us see how object indexes might help meet the Object-Dependence Challenge.

[8] The first percept is an instance of apparent motion, another common paradigm used in the literature.

[9] Or so it seems. While indexes might be applied in parallel, it is not obvious that we are able to differentiate which individual item is which, as claimed by Pylyshyn (2004). For some critical discussion regarding how to interpret multi-object tracking results, see Scholl (2009).

5.3.3 The Particulars of Object Indexes

I have suggested that the Object-Dependence Challenge is really an issue about how vision science explains object individuation. Recall that although perceptual representations are posited to explain viewpoint invariance and object specificity, explanations of the recognition *process* are silent about matters of individuation. In other words, these explanations leave it unclear how properties of identity and category are attributed to distal particulars during recognition. In contrast, research on object persistence has a clear story to tell about individuation, which requires positing object indexes. And this research points to how I can address the Object-Dependence Challenge.

I think the right response to the Object-Dependence Challenge is as follows: from the perspective of object vision more generally, the particulars to which properties are attributed are the contents of object indexes. Object indexes are a type of internal representation that is manifestly viewpoint independent. All they tell us is that there is *something* in the visual scene. So the internal representations that are indispensable to explanations of object individuation–namely, object indexes—are about distal particulars. This would seem to further suggest that they must be a kind of perceptual representation. And since the Object-Dependence Challenge is really an issue about the content of the internal representations that are posited to explain object individuation, I think the challenge can be met.

An immediate objection one might have is that there has been something of a bait and switch. My argument from object constancy is driven by how vision scientists explain the viewpoint invariance of object recognition. And as I pointed out, explanations of the recognition process are silent about how vision represents the particulars that properties of identity and category are attributed to. Granting that object indexes are viewpoint independent, to meet the challenge I would need to show that object indexes are posited in explanations *of* the recognition process. But I have not done this. I have simply shown that different lines of research on object vision posit internal representations of the right sort—of particulars and properties—but I have given no evidence that in order to explain object recognition one must posit both object representations *and* object indexes.

I have two replies to this objection. First, I would reiterate that the Dependence Challenges are supposed to be *prima facie* difficulties for arguments from constancy. And by that measure I think I have met the challenge. What I have shown is that explaining how the visual system individuates objects requires positing a type of internal representation, object indexes, which point to distal entities. It seems likely that they must be a kind of perceptual representation, and at minimum are constructed from a viewpoint independent representational scheme. Meeting the Object-Dependence challenge required a supplemental type of argument regarding object individuation, and I have provided one.

Second, I would point out that the relationship between object recognition and object persistence is simply an under-explored area of research (even the specialized field of object vision lacks unity). So I cannot rest my argument on evaluating a kind of explanation that largely does not exist. The sort of connection this objection demands has in fact been suggested in the literature. For example, Kahneman et al.

(1992, p.179) propose the following relationship between object indexes and the process of identification:

> Explicit recognition occurs at the level of which object files are currently set up. To mediate recognition, the sensory description in the object file is compared to stored representations of known objects. If and when a match is found, the identification of the object is entered in the file, together with information predicting other characteristics…

Under Kahneman, Treisman, and Gibbs' proposal, although featural properties that are bound to indexes might drive recognition, it is the target of the index to which the property is attributed. Under this story, stored object representations of identity and category are matched to those properties that are bound up with an object index—and this might include properties of momentary visual objects in a scene—but the attribution of identity or category membership is not to the momentary visual object. Rather, we can think of the attribute as a new property bound up to the object index. In which case, the property is attributed the target of the index.

This seems to be the sort of explanation the objection demands. Recall that object indexes are a kind of abstract pointer, which are assigned based on visual objects that are discriminated in a visual scene. But they are not constructed from these momentary, view-based representations. Instead they are a kind of blank slate, which are intrinsically viewpoint *independent*. When an index is assigned, we might construct a representation of surface properties, or 3D shape, but minimally an index simply tells us that there is something in the scene, and whatever it is, it is not some higher-order function of the stimulus. For while it is true that various higher-order properties might be bound to an index to help resolve the correspondence problem form moment-to-moment, this does not change the fact that these properties are not themselves the entities that indexes are directed toward.

An enticing story, but it is not one that is commonly found in research on object vision (however sensible it might be). There are object recognition results showing that spatio-temporal information plays a role (Cox et al., 2005; Tian and Grill-Spector, 2015; Vuong and Tarr, 2004; Wallis et al., 2009; Wallis and Bülthoff, 2001), and explaining these findings probably requires positing object indexes. Similarly, a different line of evidence comes from experiments on object persistence showing that object indexes can be bound with viewpoint or feature-invariant representations in long-term memory (Blaser et al., 2000; Coltheart et al., 2005; Feldman and Tremoulet, 2006; Schurgin et al., 2013) So it might be that what vision scientists *should* say is that object indexes represent the particulars for the recognition process. But, as I was careful to point out in Chapter 2, my arguments do not rest on engaging in the explanatory practices of vision science myself, but in critically evaluating them. So I will simply acknowledge that the objection points to an underdeveloped area of research in vision science, and reiterate my first reply: I have met the Object-Dependence Challenge, at least when considered as a prima facie objection.

If these replies are satisfactory, then I think it is safe to say that the Object-Dependence Challenge has been met. Now we can get to the heart of the matter, and assess whether object recognition is indeed viewpoint invariant.

5.4 Viewpoint Invariance and Stimulus-Dependence

We have seen that if object recognition is viewpoint invariant, then an argument from constancy would seem to succeed at establishing that object representations in vision are indeed perceptual representations. Such an argument easily meets the Dimension-Dependence Challenge, and as we have just seen, also meets the Object-Dependence Challenge (if we are allowed to look to research on other aspects of object vision). The question then is whether there is good evidence that object recognition *is* indeed viewpoint invariant. As it happens, a great deal of the research on the psychophysics of visual object recognition has been concerned with determining the extent of the viewpoint *dependence* of the representational schemes that underlie our capacity for object recognition.[10] Thus, I need a story for why these sort of dependency facts do not undermine my argument.

This section has two parts. First, I review some of the evidence that has been taken to show that the representational schemes that underlie recognition must be viewpoint dependent. Second, I make the case that, given the role of perceptual learning in forming and updating object representations, one must actually posit viewpoint invariant object representations *in order* to explain the evidence of viewpoint dependence. So such evidence does not make the Stimulus-Dependence Challenge a threat to my argument.

[10] A good deal of the neuroscience also provides evidence of this dependence, but I will not discuss it here (DiCarlo et al., 2012).

5.4.1 How Viewpoint Dependent is Object Recognition?

Object-centered theories of object recognition were initially developed in the absence of any psychophysical evidence for the viewpoint invariance of shape recognition (e.g., Hoffman and Richards, 1984; Marr and Nishihara, 1978). Since then, a wealth of evidence has been accrued showing that the representational schemes for objects in the visual system must be (to some extent) viewpoint dependent. Here I will review some illustrative findings.

The rationale of psychophysical experiments investigating the viewpoint dependence of object recognition is straightforward: if a representational scheme for objects is viewpoint independent, then transformations of viewpoint should have no effect on choice accuracy or reaction time (RT) when subjects perform object categorization or identification tasks. However, if transformations of viewpoint (orientation, retinal size or position, illumination) result in a change in choice and/or RTs, such that accuracy is higher and RTs are faster for some viewpoints relative to others, then this suggests that the representational scheme for objects that is being utilized for recognition is viewpoint dependent. Thus if performance is variant with respect to transformation of viewpoint, then we have evidence in favor of the viewpoint dependence of the representational schemes.

In an important study that heavily influenced subsequent research, Tarr and Pinker (1989) reasoned that the important test of viewpoint dependence was how subjects recognized novel objects when viewpoint was restricted during learning. In their study, subjects learned to recognize letter-like characters from a restricted number of viewpoints, and then tested them on viewpoints that involved transformations of orientation (picture-plane rotation). Tarr and Pinker found that both error-rate and RTs increased with difference from the training viewpoints. The same sort of pattern of results was also observed when subjects were trained to recognize 3D objects using a variety of stimuli, when novel viewpoints involve a rotation in the depth-plane (Bülthoff and Edelman, 1992; Tarr, 1995), or changes in illumination (Tarr et al., 1998a). Similar results have also been obtained for novel exemplars of familiar categories, like faces. Wallis et al. (2009) presented subjects with sequences of face stimuli that morphed between two target faces, while also undergoing transformations of viewpoint: picture-plane rotation, depth-rotation, and illumination direction. Due to the temporal association between views of the two faces, Wallis et al. found that subjects failed to notice the change in face identity. Their results suggest that object representations are constructed from associations between viewpoints that are correlated in time, and this fact can be exploited to cause errors in object identification.

Viewpoint dependence has also been observed with respect to retinotopic position. In a review of psychophysical research on the position-dependence of object recognition, Kravitz et al. (2008) observed that across multiple behavioral paradigms, performance accuracy decreased with transformations of viewpoint for even small translations of visual angle in the visual field. In one elegant study, Cox et al. (2005) had subjects saccade toward a target object in the periphery, while changing the target identity mid saccade[11] This manipulation resulted in an

[11] Saccadic eye-movements are ballistic, meaning that once an eye-movement is initiated, we lack

interesting position dependence of the subjects' representations of object identity. The result of the study was that for object stimuli that were swapped in a position-dependent manner, subjects were more likely to confuse different objects at different retinal positions as the same object, and the same object at different positions as different objects. In another striking (and puzzling) example of position-dependence, Afraz et al. (2010) observed a position-dependence of face gender discrimination, with distinct and stable patterns of position-dependence between subjects. They also observed that gender neutral faces took on a different appearance when presented within the gender biased regions of the visual field.

Recall that the Boring Principle entails that the content of a representation is that which remains invariant when its tokening is invariant. But the substantial evidence for variability in recognition performance suggests that whether and how we token object representations varies considerably with viewpoint. This would also seem to imply that the content of object representations is not object-specific. For example, there is some evidence that for both novel and familiar objects observers have preferred "canonical views": viewpoints that most readily come to mind when imagining an object, are the "best" for photographing an object, or for which observers have lowest choice error and RTs (Blanz et al., 1999). One interpretation of the viewpoint dependence of recognition is that the content of object representations is specific to visual features of canonical views (cf. Cutzu and Edelman, 1994).

While there is considerable evidence in favor of viewpoint dependence, some

the ability of changing movement trajectory. During the period of a saccade, no visual input reaches the cortex. Thus we are effectively blind—a fact that was cleverly exploited by this experiment.

have argued that there is invariant performance for object recognition under appropriately defined circumstances. Experiments investigating dependence with respect to rotations of 3D objects in the depth-plane were in part motivated to test predictions of object-centered theories like Biederman's RBC (Recognition-by-Components). Under RBC, object representations are built using collections of volumetric primitives ("geons") are that are combined into unique geon structural descriptions (GSD) of objects, and stored in long-term memory. Under Biederman's theory, one would only expect to see viewpoint invariance if three conditions are met (Biederman and Gerhardstein, 1993): (i) the shape of an object must be decomposable into GSD; (ii) the object has a unique GSD; and (iii) the same GSD is constructed for the different viewpoints. For example, a coffee cup provides a good example of an object for which (i)-(iii) should apply. It is decomposable into two geons (a cylinder and "macaroni"), has a unique GSD (though not with respect to other coffee cups), and so long as a viewpoint of the cup does not involve (self) occlusion of the handle, the same GSD will be extracted from the 2D retinal image.

While there has been some results suggesting that when (i)-(iii) are satisfied recognition performance is invariant in accuracy and RTs (Biederman and Bar, 1999; Biederman et al., 1999) other results have suggested the opposite conclusion (Hayward and Tarr, 1997; Hayward and Williams, 2000; Tarr et al., 1997, 1998b). However, even granting that there is complete performance invariance when (i)-(iii) are satisfied, such a fact cannot ground the sort of viewpoint invariance necessary for my argument from object constancy.

First, (i)-(iii) defines a limited set of possible transformations which only apply with respect to rotations in the depth-plane. To consider how limited the conditions are, consider how many objects we recognize on a regular basis involve self occlusion of parts (as a home experiment, simply look at all the things in the room in which you are currently sitting, and observe how many of them involve some level of self-occlusion). Thus, it is at best a kind of dimension-dependent invariance. Second, this sort of invariance only applies to a very restricted set of objects: they must be both identifiable in terms of GSDs, and have a unique GSD. Thus any subordinate categories that might rest on color, texture, or non-GSD shape properties are outside the scope of this invariance generalization (again, what seems like an intuitively expansive class of objects).

Perhaps the most important reason for why this sort of invariance is insufficient is that (iii) entails that the same shape feature of objects must always be visible across viewpoints. As Biederman (2000) acknowledges, one could thus take his view to specify the sort of shape features that must be present in canonical views of objects. By way of example, consider a point made by Tarr and Bülthoff (1995) that one could construct analogs of Biederman and Gerhardstein's view as a theory of color-based object recognition. Imagine an experiment where one must discriminate object categories that each have a uniquely colored feature. As Tarr and Bulthoff point out, so long as the features are visible, color provides a perfect diagnostic feature for recognizing the objects, and one would expect recognition performance to be invariant (for empirical proof of this obvious result, see Hayward and Williams, 2000). However, just as we might readily wonder whether we are doing anything but perceiving the colors (as opposed to the objects), we might likewise wonder whether (i)-(iii) really entail invariance, since the diagnostic feature must always be visible.

So I do not think that the sort of limited performance invariance that has been argued for in the literature provides any evidence against the general viewpoint dependence of object recognition. Indeed, even proponents of GSDs themselves acknowledge that recognition, under *anyone's* view, must be viewpoint dependent (Biederman, 2000; Stankiewicz, 2002). Thus it is something of a consensus that object recognition is stimulus dependent with respects to transformations of orientation (Hayward, 2003), as well as other dimensions of viewpoint.

If theories and models of object recognition posit representational schemes that must be viewpoint dependent, then this raises the question of why we should think the contents of object representations that they posit are viewpoint invariant. Indeed, some researchers have gravitated towards talk of the "tolerance" of object recognition to some transformations in viewpoint (e.g., Rust and Stocker, 2010), instead of invariance. But simply showing that recognition is tolerant to transformations of viewpoint fails to give us reason for thinking that the posited internal representations are object-specific. Thus, if my argument from constancy is to succeed, I need to show why the viewpoint dependence of the representational schemes for objects does not entail that the content of object representations is stimulus-dependence.

5.4.2 Perceptual Learning and Stimulus-Dependence

We have seen that whatever the precise representational scheme that the brain utilizes in object recognition, all theories consider it to be viewpoint dependent. At the same time, my argument depends on the content of object representations being viewpoint invariant. If viewpoint dependence entails stimulus-dependence, then my argument from object constancy fails: one cannot posit representations that are invariant with respect to viewpoint, when they are part of an representational scheme that is viewpoint dependent.

In this section, I argue that theoretically you can have viewpoint invariant content in a viewpoint dependent representational scheme. It is worth noting that this is not a novel position. Explaining how the apparent viewpoint invariant representations come from view-dependent representational schemes is at the heart of contemporary research that adopts some form of view-centered approach to object recognition (DiCarlo et al., 2012). However it is generally taken for granted that viewpoint invariance of content and viewpoint dependence of representational schemes are in fact compatible. Here I provide an argument for this assumption.

Recall that whether an argument from constancy falls prey to the Stimulus-Dependence Challenge depends on showing that, in line with the Boring Principle, the transformations across which perception are invariant involve stability in properties of the sensory input, but variability in the state of the distal world. It was for this reason that illusions do not provide conclusive evidence against an objective reading of perceptual constancies. On the one hand, the existence of an illusion suggests a lack of invariance (with respect to the distal world) across some transformations. On the other hand, if illusions are themselves perspectival (as many are), then it is still possible that an objective reading provides the best explanation of some constancy phenomenon. In short, when it comes to the Stimulus-Dependence Challenge, and objective and subjective readings of constancy phenomena, the Boring Principle cuts both ways.

The same lessons hold with respect to the viewpoint dependence and object recognition. Part of the reason visual illusions provide a prima facie basis for a subjective reading is their relative stability. For example, no matter how many times I look at a Necker Cube it does not result in me resolving the ambiguity so that I only ever adopt one of the two possible 3D interpretations of the image. In contrast, a lack of stability in response across multiple presentations would speak against an subjective reading. In particular, evidence of perceptual learning would require an objective reading of the constancy in question. Let me lay out my reasoning to this conclusion.

In his discussion of how to reconcile research on constancy achievement and constancy mechanisms, Ittelson (1951) emphasized that it is only by consideration of how perception allows us to effectively interact with the world that it makes sense to talk about objectivity in perception: "Reference to veridical distal relationships means nothing unless it means that perception and manipulation have been mutually consistent" (p.289). What I believe Ittelson recognized is that if perception is sensitive to feedback signals from the world, then this supports an objective reading. Generally the processes that underlie perceptual illusions are encapsulated from such signals: failure to perceive the illusions under slight parametric variation of the stimulus does not result in a generalization of illusory or non-illusory perception. In contrast when there is refinement of internal representations based on feedback signals—when stimulus-dependence is not stable given such feedback—then this suggests that whatever is represented, it not simply a property of the sensory stimulus. Perceptual learning is the most straightforward example of how feedback can inform vision, whether learning is supervised or unsupervised. When learning is supervised, an observer might be trained and tested to discriminate between different stimuli, and receive explicit feedback on performance. In the case of unsupervised learning, feedback is engrained in the predictive nature of visual processing. The upshot is that if apparent stimulus-dependence is modulated by perceptual learning, then this provides evidence against a subjective reading, since it suggests that whatever the object of perception is, it goes beyond the sensory signal.[12]

To illustrate, consider an early study on how perceptual learning can influence visual perception of the Ames room by Kilpatrick (1954). Kilpatrick distinguished between two kinds of perceptual learning, which he termed *reorganizational* and *formative* learning. In the case of reorganizational learning, there is increased discrimination between familiar viewing conditions, while in the case of formative learning, there is generalization to novel viewing conditions. Or as put by Kilpatrick (1954, p.363):

> Reorganizational learning shakes up and rearranges the marbles in the

[12] Incidentally, very similar reasoning is used by Dretske (1986) to argue that his informational theory of content overcomes the "distality problem", which I discuss in the next chapter.

> bag; formative learning puts new marbles in the bag, and. . . explains how most, if not all, of the marbles got in the bag in the first place.

Both kinds of learning tell us that what we are perceiving does not reduce to something sensory, since the one involves subtracting a viewpoint from among those that map to a representation, while the other involves adding (or generalizing) to novel viewing conditions. In Kilpatrick's study, subjects we allowed to view three rooms, two of which had been distorted to look identical to the rectangular third room. Initially, under monocular viewing, participants could not visually discriminate between the three rooms. During the learning phase of the experiment, subjects were positioned to observe interactions inside the rooms such as balls bouncing against the walls of the room, or lights being traced across the rooms' surfaces. During subsequent viewing, subjects were now able to reliably discriminate between the three rooms. It is clear that both kinds of perceptual learning were exhibited by the participants in the study. First, in the learning phase subjects made generalizations to novel (interactive) viewings of the target rooms. Second, this resulted in reorganizational learning, as exhibited by their subsequent ability to visually discriminate the rooms under monocular viewing.

Under the Boring Principle the object of perception is that which is invariant when perception is invariant. If some constancy mechanism admits of both reorganizational and formative learning, this tells us that the content must be something beyond the sensory input. The fact that perceptual learning altered the perception of Kilpatrick's distorted rooms suggests that the illusory perception of such rooms does not necessarily constitute evidence against an objective reading of constancies involved with perceiving shape (with respect to the spaces).

Object representations are also heavily influenced by both reorganizational and formative learning. For one, virtually all of the experiments testing for viewpoint dependence utilize formative learning paradigms, since they involve training and then testing on novel viewpoints. In some of these studies, learning effects are clearly observed, like reduction in RTs with repeated presentations of novel viewpoints (Edelman and Bülthoff, 1992). For example, Tarr and Gauthier (1998) found that observers could generalize to novel viewpoints of an object when they had observed previously observed members of the same category from that viewpoint. And, as one would expect, when subjects were trained to become "experts" at recognizing novel objects, both error rates and RTs decreased with experience.

In short, viewpoint dependence is not stable across presentations, in which case it cannot provide evidence against the viewpoint invariance of object representations in vision. So I do not think the viewpoint dependence of object recognition performance presents a problem for my argument from object constancy. In contrast, the fact that observers can generalize to novel viewpoints, and make increasingly fine discrimination between familiar viewpoints, gives *another* reason for thinking that the contents of object representations must be viewpoint invariant, even if the representational schemes are themselves viewpoint dependent. In which case, research on viewpoint dependence provides further reason for thinking that one must posit perceptual representations to explain viewpoint invariance, and actually strengthens my argument from object constancy.

Before continuing, it is also worth considering that the extent of viewpoint dependence in recognition is likely overstated, due to the poor ecological validity of typical object recognition experiments that include manipulations of viewpoint. Under natural viewing conditions, we do not perceive objects via singular and brief snapshots from stable vantage points. Rather, transformations of viewpoint are continuous as we navigate and interact with the world. Thus one might legitimately wonder whether some of the evidence for viewpoint dependence tells us more about the demands of a task subjects are asked to perform, than the structure of the representational schemes that underlie object recognition (Schyns, 1998).

To illustrate this last point, consider an experiment by Tian and Grill-Spector (2015) that investigated the influence of unsupervised learning on object recognition. In the experiments, subjects were tasked with recognizing novel objects across transformations of orientation. During a learning phase, subjects passively viewed either sequential or random series of images of the objects. The sequential condition was akin to seeing snap shots of an object rotating in space, while the random condition involved the same snap shots randomly ordered. When presented with 24 views of the objects, subjects showed massive improvement in recognition across transformations of orientation in both the depth and picture planes. This improvement was observed after learning in both the sequential and random conditions. When reduced to just 7 views during learning, high level of improvement was still observed for the sequential condition. Tian and Grill-Spector's stimuli more closely mimic how we learn to recognize objects, via the presentation of (and continuous transformation between) series of viewpoints. Were the experiment replicated using some of the stimuli from early studies reporting viewpoint dependence effects of choice and RT, we would likely observe similar results.

5.5 Taking Stock of the Other Ingredients

In Chapter 2 I specified a number of staple and special ingredients for perceptual representation. I have been operating under the assumption that ingredients required for internal representation are satisfied, in general, by the representational schemes posited by research within the information-processing framework in vision science (as I argued in Chapter 3). The elusive ingredients for perceptual representation were perceptual objectivity and robustness. Based on the argument from object constancy that I have defended in this chapter, I think that they are possessed by the internal representations that are posited to explain how we visually recognize objects. Given my argument from object constancy, one might now wonder whether one should have granted that the other ingredients are indeed satisfied. Hence, in this section, I will look at some of the other ingredients for perceptual representation, and why they are also satisfied by the object representations posited to explain recognition. I will also finally return to discussion of the Job Description Challenge.

5.5.1 Recognition and Varieties of Robustness

I have focused on perceptual robustness, since it was this kind of robustness that was necessary for perceptual representation. The kind of robustness that I have referred to as “resilience”, or robustness with respect to intervening causes, is also necessary. In Chapter 3 I argued that in general, it is a safe assumption that representational schemes posited by the information-processing framework are presumed to be resilient. I believe this is also true when it comes to the internal representations posited to explain object recognition.

Consider, for example, the study of Afraz et al. (2006), who administered micro-stimulation to clusters of face selective neurons in the inferior temporal cortex of macaque monkeys, a region strongly associated with object representation. They found that administering the stimulation strongly biased the monkeys to make a face response, in a face/non-face recognition task. The rationale of the experiment is clearly dependent on the idea that the face-selective cells are in part constitutive of the neural implementation for object representations in the monkey visual system. If micro-stimulation was thought to alter the content of the representations, the experiment would be deeply flawed. Furthermore, the fact that a bias is observed in the response of the monkey supports the idea that the neurons are partially constitutive (in some way) of object representations for faces. Of course, over-time, repeat use of micro-stimulation might cause a learning effect, so that the monkeys no longer trusted tokenings of their representations of faces as reliable (at least, under the conditions of the experiment). But this would not be an instance of a change in the content, but rather an instance of the reorganizational learning discussed earlier. So I think we have good reason to believe that the representational schemes must be resilient.

What about robustness with respective to tokenings caused by other mental processes? In Chapter 2 I referred to this as "cognitive robustness", which I suggested was not necessary for perceptual representation. Still, one might wonder whether object representations must have this other ingredient in order to explain facts about recognition. There is an ongoing debate as to whether the internal representations utilized in recognition are the same as those recruited for mental imagery and visual simulation, such as tasks which involve mental rotation of imagined objects. Tarr and Pinker (1989) proposed that recognition and imagery rest on the same representations, which would suggest that the internal representations they posited are cognitively robust. Based on fMRI results, Gauthier et al. (2002) suggested that object recognition and mental rotation process are localized to distinct streams in the visual system, which would seem to suggest Tarr and Pinker were wrong about there being a shared encoding scheme. While it is true that imagery and simulation in mental rotation tasks recruit several other areas not involved in recognition, the inferior temporal cortex is also recruited (Zacks, 2008), which runs counter to the results of Gauthier et al. So it is at present an open question whether object representations for recognition are thought to be cognitively robust.[13]

[13] This isn't to deny that recognition and mental rotation exploit share resources—even if the two processes can be dissociated (Cheung et al., 2009). But what I think is unclear is whether the sorts of internal representations for object identity and category that are recruited during recognition are among these resources.

5.5.2 Recognition and Representational Function

Another necessary ingredient for perceptual representation is having an appropriate representational function. In Chapter 2 I identified such a function as, in general, that of being a perceptually robust stand-in. Having such a representational function—which is tied to properties of intentional content—was also necessary for meeting Ramsey's Job Description Challenge. Recall that Ramsey (2007, p.27) laid out the demands of a job description as follows.

> There needs to be some unique role or set of causal relations that warrants our saying some structure or state serves a representational function. These roles and relations should enable us to distinguish the representational from the non-representational and should provide us with conditions that delineate the sort of job representations perform, *qua* representations, in the physical system.

It seems safe to say that object representations that are posited to explain recognition serve as perceptually robust stand-ins, and can therefore meet Ramsey's challenge. First, the functional significance of object representations, under anyone's view, conforms generally with that posited by Marr and Nishihara, which is in terms of a stand-in function. Second, one can find explicit evidence that internal states that function as *perceptually robust* stand-ins are posited to explain recognition, in so far as they function as stand-ins in the face of transformations of viewpoint. For example, here is how DiCarlo et al. (2012, p.416) define the information-processing task of recognition:

> . . . we and others define object recognition as the ability to assign labels (e.g., nouns) to particular objects, ranging from precise labels ("identification") to course labels ("categorization"). More specifically, we focus on the ability to complete such tasks over a range of identity preserving transformations (e.g., changes in object position, size, pose, and background context). . .

To explain recognition, internal representations are posited that have content that is viewpoint invariance and object specific. In light of an objective reading of this invariance, and my argument from object constancy, this entails that the content of these representations must also be perceptually robust. So they must be viewpoint invariant stand-ins, which entails they must also be perceptually robust stand-ins as well. So given my argument from object constancy, and the fact that object representations must function as perceptually robust stand-ins for object identity and category, I think there can be little doubt that these posits meet Ramsey's Job Description Challenge.

5.6 Conclusion

We have, then, an argument for my second main conclusion. To explain facts about object recognition, internal representations are posited with content that is both viewpoint invariant and object specific. And since these "object representations" have all the ingredients for perceptual representation, a notion of perceptual representation is therefore indispensable to the explanations.

So we have seen why a notion of perceptual representation is indispensable to certain explanatory practices in vision science. Without such a notion, one would fail to adequately explain the phenomenon of interest. What I will now show is that typical explanations of object recognition rest on problematic assumptions about how representational content is fixed, which are endemic to research within the information-processing framework of vision science. Informational theories of content, I believe, offer hope of a cure.

Chapter 6: Representation Transformed?

6.1 Introduction

We have now seen why a notion of perceptual representation is indeed indispensable to theories and models aimed at explaining facts about visual object recognition; in particular, the viewpoint invariance and object specificity of object representations. my argument for this conclusion amounts to a defense of the Distal Object Thesis as a theme in the story of vision:

> *The Distal Object Thesis*: the end-stage internal representations that underlie object vision track properties of entities in the distal world.

The other main theme of interest, recall, was the Transformational Thesis:

> *The Transformational Thesis*: the function of the visual system is to reformat ("transform") information latent in the retinal image into a representational format that makes it available for use by further perceptual or cognitive systems.

In this final chapter I return to the topic that I began with: how the Transformational Thesis is in tension with the Distal Object Thesis, and why informational theories of content provide strategies for relieving this tension.

In this chapter I do three things. First, I argue that the Transformational Thesis rests on the presumption that representational content is fixed by a reliable, information carrying causal relation between a representation and what it is about. In other words, these explanations tacitly assume an informational theory of content. Since explanations of object recognition tend to depend on the Transformational Thesis, they make the same assumption. Second, I argue that the success of explanations of object recognition depends on a solution to the disjunction problem. At its core, the disjunction problem shows why informational theories cannot account for the robustness of intentional content (Fodor, 1990). Explanations of object recognition attempt to characterize the reliable causal mechanisms that produce representations that are both viewpoint invariance and object specific. As argued in the last chapter, the content of an object representation has these properties only if it is perceptually robust. If content is fixed solely by a reliable causal relation, then it is not perceptually robust, and it cannot have these ingredients for intentional content. In which case, if explanations of object recognition are to succeed, the informational theory they assume must be supplemented. Third, I very briefly review how some of the usual (suspect) solutions to the disjunction problem—which appeal to learning (Dretske, 1981), teleological function (Dretske, 1988; Neander, 2012), or counterfactuals (Fodor, 1987, 1990)—might be paired with explanations of object recognition. Although each of these purported solutions faces problems that put them in tension with the story of vision, the general upshot of my discussion is that informational theories do provide promising strategies for resolving a tension in the information-processing narrative of vision science.

The rest of the chapter is structured as follows. In Subsection 2, I describe the structure of informational theories in more detail, and their shared origins with the information-processing framework in vision science. I also clarify the connection between the disjunction problem and robustness. In Subsection 3, I provide evidence that an informational theory is a presupposition of the information-processing framework in general (via the Transformational Thesis), and of explanations of visual object recognition in particular. I further argue that for this reason, the success of explanations of viewpoint invariance and object recognition depends on a solution to the disjunction problem. Thus, any theory of content that solves the problem is of relevance to these explanations—which was Conclusion 3 of my project. In Subsection 4, I briefly review the usual proposals for solving the disjunction problem. I evaluate these solutions with respect to how well they cohere with the facts of object recognition, and the story of vision. In Subsection 5, I provide a summary conclusion of this chapter, and the project as a whole.

6.2 The Disjunction Problem for Informational Theories of Content

In this section I describe informational theories, and their disjunction problem, in a bit more detail. I emphasize two things. First, informational theories are in part motivated by the organization of information-processing systems. Second, the disjunction problem is at its core a problem about how informational theories are supposed to account for robustness—as opposed to how they are supposed to account for error (as is sometimes claimed).

6.2.1 The Character of Informational Theories

What all informational theories have in common is the idea that the following provides at least a constitutive condition on how intentional content is fixed, for any mental representation:

> (Inf) There is a reliable, information carrying relationship between the representation and its content.

The sense of "information" here is that of what has been termed *natural* information (Dretske, 1988), which I introduced in Chapter 3. To rehearse the examples from earlier, this is the sort of information that tree rings carry about the age of a tree, smoke carries about fire, the whistle of a kettle carries about water temperature, mercury in a thermometer carries about ambient temperature, and a compass needle carries about the direction of the nearest magnetic pole. In each of these cases, due to the dependence of one event on another event, the former is informative about the occurrence of the latter. While there are many different proposals about how we are to understand the notion of natural information, at least when it comes to intentional content, the usual assumption is that it is a kind of nomic, causal dependence (Dretske, 1988; Fodor, 1990). This is the characterization I have been assuming so far, and I will continue to assume it in what follows.[1]

What is often passed over in discussions of informational theories is the initial motivation for the view. Why is carrying information a *plausible* constitutive con-

[1] Hence, what I am terming informational theories are also often called "causal", or "causal-informational" theories of content (Rupert, 2008).

dition for intentional content? Informational theories were developed, in part, with the vertical aim of naturalizing intentional content, and hence the mind. However, at least initially, they were also supposed to: "make sense of. . . the theoretically central role information plays in the descriptive and explanatory efforts of cognitive scientists. . . " (Dretske, 1983, p.55). And it was in part reflection on this role that provided the initial motivation for informational theories.

The backdrop for informational theories of content is a commitment to RTM. Often going hand-in-hand with RTM is the empirical hypothesis that the brain is some kind of representational, or information-processing system (Pylyshyn, 1984). If the brain is such a system, then neural states carry and process natural information in a manner that is similar to artificial systems. And when it comes to the sorts of information-processing devices that humans engineer, their internal states have their content by carrying natural information about the signals that the devices process.[2]

Consider some simple (non-information-processing) artifacts: the whistling kettle, mercury thermometer, and magnetic compass. The reason the whistle of the kettle, height of the mercury, and position of the compass needle are (respectively) informative of water temperature, ambient temperature, and the cardinal directions is because they were designed to exploit certain nomic dependencies. One thought that motivated the development of informational theories is that the brain is also built so that there are nomic dependencies between its internal states and states of

[2] Here the terminological assumptions about "information-processing" in Chapter 3 still apply: by "information-processing" I have in mind the processing of natural information.

the distal world. Assuming that we can discharge the need of a designing *agent*, then we have a naturalistic basis for content. All we need to do is simply *identify* the content of a internal state with the information it carries. (Dretske, 1988, p.54) makes the same point using the example of a weight scale:

> Although a great deal of intelligent thought and purpose went into the design and manufacture of an ordinary bathroom scale, once the scale has been finished and placed into use there is nothing conventional, purposeful, or intelligent about its operation. This device indicates what it does without any cooperation or help from either its maker or its user. All you do is get *on* it. It then gives you the bad news.

Let me give two examples of this sort of reasoning—from natural information-processing in devices, to the fixing of representational content—in action.

1. Dretske (1981) was the first to explicitly propose an informational theory of intentional content (there are antecedents in Stampe, 1977), and took as his inspiration Shannon's theory of communication (Shannon, 1948). Crucially, information in Shannon's sense is not an *semantic* notion. As a probabilistic measure of uncertainty it does not require there to be a message in the sense that something is in fact being communicated between a "sender" and "receiver". However, Dretske believed that Shannon's work brought into focus what does determine the content of a signal:

> . . . though [information] theory has its attention elsewhere, it does. . . highlight the relevant objective relations on which the communication of genuine

> information depends. For what this theory tells us is that the amount of information at [a receiver] about [a source] is a function of the *degree of lawful (nomic) dependence* between conditions at these two points. If two conditions are statistically independent (the way the ringing of *your* telephone is independent of the ringing of *mine*), then the one event carries no information about the other. When there is a lawful regularity between two events, statistical or otherwise, as there is between your dialing my number and my phone's ringing, then we speak of one event's [sic] carrying information about the other. And, of course, this is the way we do speak. (Dretske, 1983, p.56 emphasis in original)

Even though Shannon was not interested in questions relating to the quality of information, Dretske believed his theory highlighted the fundamental sort of dependence that establishes what a signal is about. Fundamental to Shannon's theory are conditional probabilities between different events, which Dretske interpreted as measures of the degree of dependency between the events. In which case, we need only identify the content of a signal with the events it is dependent on:

> ...the quantities of interest to [information] theory are statistical functions of these probabilities. It is this *presupposed* idea that I exploit to develop an account of a signal's content. These conditional probabilities determine how much, and indirectly what, information a particular signal carries about a remote source. One needs only to stipulate that the content of the signal, the information it carries, be expressed by a

> sentence describing the condition (at the source) on which the signal depends in some regular, lawful way. (Dretske, 1983, p.57 emphasis in original)

Dropping the requirement that the content being described in sentential terms, much the same idea is found in other probabilistic approaches to natural information (Scarantino, 2015; Skyrms, 2010), which take Dretske, and Shannon's communication theory, as their starting points. Under these other proposals it is still the case that some information-theoretic measure of the quantity of natural information can be used to identify the content of an informational signal based on the strength of the dependency relation between a receiver and a signal. Although it has been common to dispense with a probabilistic account of natural information, what should be clear is that informational theories of content were derived from thinking about how the content of internal states of an information-processing system is determined.

2. Similar reasoning to Dretske's is also found in Enç (1982), who was interested in RTM, and providing a functional analysis of psychological states. Here is what he says about the content of internal states of some information-processing system:

> Suppose now we shift our terminology and speak of the *content* of a functional state of a system as being constituted by the specific construction of the properties of the event which the system has the functioning of responding to and as result of which the system, under the conditions of normal function, enters that state. In other words, the content of a

> functional state will be determined by certain properties of the event that causes the system to enter that state... The appropriate specific construction of these properties will be determined by the nature of the mechanisms by means of which the system, when it is well functioning, processes the information about these properties that is contained in its internal representation; in other words, it will be determined by the correct aetiological account of how the system normally enters the functional state. (Enç, 1982, p.175-176 emphasis in original)

Enc makes explicit reference to "normal functioning"—a topic we will return to—which raises other issues as we shall see below, but what should be clear is that Enc, like Dretske, is making an inference from how content is determined in information-processing systems. Like Dretske, Enc proposes we now apply the same picture to human psychology, for which we give a functional treatment:

> ... analogous to the way the contents of the functional states of physical systems are determined, the contents of psychological states will be determined by the specific construction of the relevant properties that are involved in the aetiology of these states. (Enç, 1982, p.180)

Here Enc draws the same conclusion as Dretske: content is fixed in a manner analogous to other information-processing systems. So we have now seen that the motivation for an informational theory comes from reflecting on how content is fixed in other information-processing systems. We now turn to the primary challenge for informational theories.

6.2.2 The Disjunctive Difficulty of Robustness

The disjunction problem was initially proposed by Fodor (1984), who saw it as a fundamental challenge to all informational theories of content—though Dretske (1981) seemingly anticipated the difficulty. There are of course many objections that have been brought against informational theories. Some of these are idiosyncratic to particular theories, while others apply to informational theories in general.[3] Some of these might also present difficulties for explanations within the information-processing framework that presume an informational theory. For example, one issue we shall touch on later is how informational theories are to allow for the representation of distal causes—a problem arguably as relevant as the disjunction problem (cf. Godfrey-Smith, 1989). Still the disjunction problem is widely regarded as perhaps the most fundamental challenge to informational theories. Hence I will set aside other potentially relevant objections to informational theories.

Let us remind ourselves of the structure of the problem. As we have seen, some tokenings of a representation are *wild* in the sense that they are caused by something other than what the state represents. So the causal dependence between a representation and its content is imperfect (Fodor, 1984, p.240). But if the content of a representation is determined simply by whatever reliably causes the state to be tokened, then the representation will be about the disjunction of its reliable causes. The problem is traditionally illustrated using instances of misrepresentation. Recall

[3] There are several reviews that cover informational theories, and their difficulties (Adams and Aizawa, 2010; Cohen, 2004; Loewer, 1987; Rupert, 2008).

Raccoon from Chapter 2. When walking down a back alley on a dark night I might mistake a raccoon (or as is more likely in Maryland, a possum) for a cat. In this situation, I misrepresent the world, as my concept CAT has been caused by something other than a cat. Misrepresenting is an instance of a wild tokening of a representation. Even though raccoons or possums might be reliably mistaken for cats under degraded viewing conditions, it is implausible that this dependency entails that my concept CAT also represents these animals.

Misrepresentation is a fact about how we represent the world. A "crude" informational theory cannot capture this fact. If reliable causation is taken as sufficient for determining content, cat-ish looking raccoons and possums are indeed part of the disjunctive content of my concept CAT. Put differently, a crude informational theory gets the content ascription in the case of misrepresentation wrong. This descriptive inadequacy brings out an important difference between the sort of systems and devices that inspired Dretske to develop an informational theory. Thermometers and compasses register information about temperature and polar direction, and are quite reliable in carrying out these functions—assuming appropriate environmental conditions. Change the altitude of the thermometer, and it will not accurately gauge ambient temperature. Sail close to the equator, and a magnetic compass will vacillate between the directions of the magnetic poles. In these situations, the devices make no "errors", and indeed they seem incapable of doing so—though we might make plenty of mistakes, if we are careless enough to use them (Dretske, 1986). So while all informational theory holds that intentional content is fixed in a way that relies on the same naturalistic ingredients as these devices, further ingredients are clearly required.

While the disjunction problem strikes at the heart of informational theories, there is persistent confusion in the literature with respect to what the problem is really about. Part of this stems from the fact that the problem is typically illustrated using examples of misrepresentation, which has resulted in it being characterized as a facet of the "problem of error" for theories of content—the purported requirement that theories of intentional content must account for, or explain, how representations can be mistaken. Thus under what we may call the *erroneous interpretation* of the disjunction problem, it is simply an aspect (or alternative characterization) of the problem of error.

The erroneous interpretation errs in two ways. The first harkens back to some of the discussion from Chapter 2, where I tried to distinguish questions of intentional content from those of representational function. Many philosophers claim that all theories of content must explain how misrepresentation is possible, but explaining how representational errors can occur is arguably a task for an account of representational function, not of intentional content (Cummins, 1996; Fodor, 1990).[4] The second is to presume that the disjunction problem thus simply illustrates the problem of error for informational theories. However, this is to have matters in reverse. Rather instances of error such as misrepresentation instead are illustrative of what the disjunction problem is really about: robustness. The point is made

[4] For some instances of this "error about error" see: Cohen (2004, p.216), Dretske (1988, p.65), Godfrey-Smith (1989, p.537), Neander (012b), Ramsey (2007, p.129), and Rupert (2008, p.356).

clearly by Fodor (1990, p.90 emphasis in original):[5]

> Errors raise the disjunction problem, but the disjunction problem isn't really, deep down, a problem about error. What the disjunction problem is really about deep down is the difference between *meaning* and *information*... Information is tied to etiology in a way that meaning isn't. If the tokens of a symbol have two kinds of etiologies, it follows that there are two kinds of information that tokens of that symbol carry... By contrast, *the meaning of a symbol is one of the things that all of its tokenings have in common, however they may happen to be caused*... So, information follows etiology and meaning doesn't, and that's why you get a disjunction problem if you identify the meaning of a symbol with the information that its tokens carry. Error is merely illustrative; it comes into the disjunction problem only because it's so plausible that the false tokens of a symbol have a different kind of causal history (and hence carry different information) than the true ones.

So the disjunction problem is not about error. Rather, it is about:

> what one might call the *robustness* of meaning... Solving the disjunction problem and making clear how a symbol's meaning could be so insensitive to variability in the causes of its tokenings are really two ways of

[5] An interpretive note: Fodor uses 'meaning' and 'content' more or less interchangeably. As I outlined in Chapter 2, when talking of intentional content I have in mind the distal referent of a mental representation.

> describing the same undertaking. If there's going to be a causal theory of content, there has to be some way of picking out *semantically relevant* causal relations from all the other kinds of causal relations that the tokens of a symbol can enter into. (Fodor, 1990, p.91 emphasis in original)

All informational theories hold that a reliable, information carrying causal dependency obtains between a representation and its content. This is constitutive of the representing relation. However not all such dependencies are content determining. In this respect, intentional content is robust. As discussed in Chapter 2, misrepresentation constitutes an instance of perceptual robustness, but there are other kinds of robustness as well, including cognitive robustness (e.g., when thinking causes a token of CAT when none are present) and resilience (e.g., when knock on the head causes me to token CAT). In these cases, there is no misrepresentation, but the disjunction problem can be raised just the same (Fodor, 990b; Rey, 1998). Let us call this the *robust interpretation* of the disjunction problem.

I believe the robust interpretation provides the correct characterization of the disjunction problem. A solution to the disjunction problem should allow for the possibility of error, as exhibited by instances of misrepresentation, but what it must *explain* is the relevant forms of robustness. This is significant since we have already seen that perceptual robustness is closely related to the explananda of research on visual object recognition. Thus, for present purposes we may focus on the disjunction problem specifically as it relates to perceptual robustness. The question then is how (Inf) from earlier can describe a content determining relation, but the following does not:

> (Mis) There is a reliable, information carrying relation between the representation and "wild" events.

Where (Mis) holds for any mental representation, and some set of "wild"' events on which the state nomically depends, but which are not part of what it represents.

We have now seen why the disjunction problem presents a fundamental challenge to informational theories. We have also gotten clearer on what the problem is really about. What I now wish to show is that the problem also presents a fundamental challenge to research within the information-processing framework in vision science that relies on the Transformational Thesis, including research on object recognition.

6.3 The Disjunction Problem and Vision Science

There is the perception among some philosophers that the disjunction problem is a quaint reminder of the trouble you can get into when one becomes too invested in trying to naturalize the mind. The problem does not reflect, and has no bearing on, research within cognitive science. For example, here is Burge's discussion—or rather, dismissal—of the problem:

> A problem that has exercised those who try to find reductive explanations of the notion of representation is called the *Disjunction Prob-*

> *lem*. . . The challenge is to explain conditions on representation that show why representations represent one range of entities rather than other entities that co-vary with, and in many cases play a role in causing, the representation. . . [Dretske and Fodor] pose the problem with no reference to specific empirical work in psychology. Their versions of the problem are correspondingly artificial. The Disjunction Problem is largely an artifact of reductive programs, detached from explanations in perceptual psychology. . . (Burge, 2010, p.322)

I believe Burge, and others who share his sentiment, are mistaken, at least with respect to vision science. The disjunction problem is not an "artifact" of the project of naturalization, and has real consequences for the explanations of visual object recognition.

In this section I first provide evidence that an informational theory of content is a general presumption of the information-processing framework in vision science, via the Transformational Thesis. This tacit assumption of an informational theory puts explanations that rely on the Transformational Thesis at risk of the disjunction problem. I go on to show why the problem therefore poses a threat to explanations of object recognition.

6.3.1 Informational Theories and the Transformational Thesis

It is easy to find cases where an informational theory is explicitly identified as a core assumption of the information-processing framework in vision science. For example, in his influential textbook, Palmer (1999, p.78) justifies the assumption as follows:

> The causal factor…is important for two reasons. One is that for the representation to be current, as a perceptual representation must be, it requires constant updating. A causal chain from events in the external world to events in the internal representation is an ideal way (though not the only way) to achieve this. The other is that for the representation to be authentic, rather than accidental, there must be some linkage to the world it represents. Again, a causal connection seems to be the ideal solution.

The second reason provided by Palmer is clearly connected to how the content of perceptual representations is fixed: to represent the world a state must be connected to the world, and causation provides this connection. Pylyshyn (2007) also identifies an informational theory as an assumption of vision science, while pointing out that it provides a starting point, rather than a complete picture, of how content is fixed:

> The implicit understanding is that what representations represent is in some way traceable to what caused them, or at least what might have caused them in a typical setting…This is certainly a reasonable starting assumption, but it is incomplete in crucial ways; there are generally very many ways that any particular representation could have been caused,

> yet the representation may nonetheless unambiguously represent just one scene. (Pylyshyn, 2007, p.5)

As Pylyshyn is careful to point out, carrying information does not provide a sufficient condition because there is a: "gap between the incoming causally linked information and *representational content*" (Pylyshyn, 2007, p.3 emphasis in original), when it comes to the internal states of the visual system. To the extent that these passages can be taken as representative of the prevailing viewpoint within the information-processing framework, they show that an informational theory is often a (tacit) assumption of the framework.[6] But what I need to now show is why the Transformational Thesis presupposes something close to a *crude* informational theory, according to which carrying information is taken to be sufficient for determining content.

The Transformational Thesis amounts to the claims that all the needed natural information about a stimulus is latent in the retinal image, and the function of the visual system is to reformat, or "transform" this information into a format that

[6] Though I do not think that it is the only theory of content that is assumed. For example, Palmer (1999) also appeals to similarity relations between an internal representation and what it is about, in line with the notion of s-representation from Chapter 2. Also Shagrir (2010) has argued that a notion of s-representation is tacit in Marr (1982). And it is not implausible that the notion of encoding and decoding information suggests some sort of "two-factor" approach, according to which content is determined by both information, and conceptual role. But my goal here is not to provide a complete account of the assumptions that vision scientists make about how content is fixed, but rather just those related to the Transformational Thesis. Those who deny the thesis may not even be committed to an informational theory at all.

makes the information available for use by further perceptual or cognitive systems. I will now provide an illustration of the importance of the Transformational Thesis to explanations within the information-processing framework, and why it appears to depend on a commitment to a crude informational theory.

The Transformational Thesis has been quite central to recent methodological developments in the cognitive neuroscience of vision. Increasingly researchers are using multi-variate pattern analysis, or "decoding", methods to determine what sort of information about experimental conditions are latent in measured patterns of brain activity. In this research, machine learning classifiers are trained to discriminate patterns of neural activity (from cellular recordings, or non-invasive techniques such as fMRI or EEG/MEG) from different experimental conditions. If the classifier is able to perform above chance when assigning labels to activity patterns, then this minimally shows that (natural) information about the conditions is latent in the patterns.[7] However, decoding results are typically pitched as revealing the *content* of the representations implemented in the brain region that produces the measured patterns of activity. For example, in a discussion of decoding methods in fMRI research, Mur et al. (2009, p.1 emphasis in original) draw a contrast between traditional "activation-based" methods—which look for signal amplitude changes in the hemodynamic response—and "information-based" approaches:

> Activation-based analysis aims to detect regional-average activation dif-

[7] This description is quite cursory. For an introduction to decoding methods for fMRI see Norman et al. (2006). A more technical introduction to the machine learning techniques they depend on is provided by Pereira et al. (2009).

> ferences and infer *involvement* of the region in a specific mental function. Pattern-information analysis, by contrast, aims to detect activity-pattern differences and infer *representational content*.

Norman et al. (2006, p.425) offer a very similar theoretical rationale for decoding methods:

> The MVPA approach assumes that cognitive states consist of multiple aspects ('dimensions'), and that different values along a particular dimension are represented by different patterns of neural firing. This implies that we can measure how strongly cognitive dimension x is represented in brain region y, by measuring how much the pattern of neural activity in region y changes, as a function of changes along dimension x. Here we are using 'region y represents dimension x' to mean 'region y carries information about dimension x'...

To further drive the point home, here is how Kriegeskorte and Kievit (2013, p.402) define "explicit representation" within the context of decoding research:

> [as a] neuronal representation of a stimulus property that allows immediate readout of the property by downstream neurons. If the property can be read out by means of a linear combination of the activities of the neurons...the property is explicitly represented.

So the inference relied on by decoding research appears to be as follows, when it comes to evidence of above chance decoding from some set of neural activation patterns:

> (D1) Natural information about experimental conditions is latent in some pattern of neural activity.
>
> (D2) Therefore, we have good evidence that the patterns implement a internal *representation* of the experimental conditions.

A good deal of the most important work using decoding methods has been on visual object recognition (e.g., Haxby et al., 2001; Kriegeskorte et al., 2008), so it is safe to assume that this sort of inference is used to draw conclusions about object representations (a point to which I will return). Clearly the one would only infer (D2) from (D1) if one thinks that carrying information is sufficient for determining content. So given our discussion of informational theories, it is a bad inference.[8] But (D2) follows quite naturally if one endorses the Transformational Thesis. As pointed out by Cox (2014, p.189), decoding research:

> implicitly recognizes that the problem of vision is not one of information content, but of format. We know that the activity of retinal ganglion cells contains all of the information that the visual system can act upon, and that nonlinearity and noise in neuronal processing can only decrease (and never increase) the absolute amount of information present. However, the information present in the firing of retinal ganglion cells is not in a format that can be easily read-out by a downstream neuron in order to guide action.

[8] For further discussion of the problematic role this inference plays in decoding research, see Ritchie et al. (shed).

So it is *part* of the Transformational Thesis that carrying natural information suffices for determining content (when it is appropriately formatted). It is worth emphasizing that the thesis, and specifically, the idea that all the information needed for vision is "in" the retinal image, has been a reoccurring theme within the information-processing framework since its early beginnings. The root of the Transformational Thesis is often associated with Gibson (1950, 1979), who held that all the information the observer can act upon is available in the retinal image. While Gibson sought to move from this claim to reject representation and process in vision, it has been a common move, as exemplified by the passage from Cox, to grant that the information is present, but not appropriately formatted (hence the need for further internal representation and information-processing).

That research in the information-processing framework shares this common point of departure with Gibson has been frequently observed. For example, the neo-Gibsonians Withagen and Chemero (2009, p.382 n.2) claim that Marr: "argued that Gibson was right about the variables that animals rely on [in visual perception], but argued that the detection of specifying variables still requires computational/inferential processes." In fact, this common point of departure with Gibson goes back to the birth of the information-processing framework, which occurred with the first application of Shannon's communication theory to visual perception. Attneave (1954) hypothesized that the function of vision was to reduce informational redundancy.[9] Whatever the merits of this idea (see Barlow, 2001, for a critical ret-

[9] Roughly, the idea was that different portions of the visual input are highly predictive of each other. Thus, it should be possible to form a more compressed formatting of the visual input

rospective), it rested on the claim that the challenge for vision is one of formatting (i.e., reducing redundancy). For example, in a review of early approaches to perceptual psychology Bevan (1958, p.42) says the following of Attneave's then novel approach:

> This view, like Gibson's, assumes an isomorphism between the spatial properties of the external object and the organism's perception. . . The theory thus implies an intact, completely developed scanner, the character of the perception reflecting essentially the character of the input.

In other words, Bevan is suggesting that, under Attneave's view, the necessary information for perception is there in the retinal image. Interestingly, even Dretske (1983, p.62 n.1) acknowledges that his claim that the retinal input carries information about a stimulus is close to the views of Gibson (1950). In fact, it is not hard to see the same sort of considerations that motivated Dretske (1981) in developing an informational theory at play when it comes to the Transformational Thesis. Namely, it only makes sense to say that the function of the visual system is to (merely) re-format the information made available in the retinal image if we identify the content of a representation with the natural information that it carries. Hence, again we see how an informational theory has fallen out of thinking of some aspect of the mind in information-processing terms.

The foregoing suggests that, in general, the information-processing framework requires some solution to the disjunction problem. But so far we have not seen

without loss of (Shannon) information.

what sort of explanatory damage the problem can cause, absent a solution. Spelling out potential for such damage, with respect to explanations of object recognition, is what I turn to next.

6.3.2 The Problem and Explanations of Object Recognition

My argument from object constancy in the last chapter roughly ran as follows: given the factual claim that the content of object representations is viewpoint invariant, then it follows that it is perceptually objective. Given the further factual claims that inputs to the retina are underdetermining and variable, then it also follows that the content is perceptually robust. Thus object representations posited to explain recognition must possess the two key special ingredients for intentional content of perceptual representations. It also follows, given the structure of my argument from object constancy, that there is viewpoint invariance only if there is perceptual robustness.

However if the content of some internal representation is determined by a reliable, information carrying causal relation—and not much else—then the content cannot be perceptually robust, and so the representation cannot be a perceptual representation. This is the consequence of the disjunction problem. We have now seen that the information-processing framework seems to presuppose something like a crude informational theory, as evinced by the importance of the Transformational Thesis within the research program. What consequences does this now all have for explanations of object recognition, and the Distal Object Thesis?

So far I have argued that the primary explananda for research on object recognition are facts about the content of object representations: their viewpoint invariance and object specificity. The sorts of explanations that are on offer pay homage to the Transformational Thesis: they are attempts to explain the reliable, mechanistic process by which object representations are constructed, which have these properties. But such a mechanism that establishes a reliable, information carrying relation is not sufficient for content that is perceptually robust. Since object representations are viewpoint invariant and object specific only if they are perceptually robust, then such a mechanism cannot produce representations that have contents with these properties. So the upshot is that by (tacitly) assuming an informational theory that falls prey to the disjunction problem, the very structure of explanations of object recognition entails that these explanations cannot succeed.

To make this all a bit more concrete, recall the general structure of explanations of object recognition from the last chapter: an internal representation (a structural description of an object's 3D shape, or a view-centered representation) is matched to a set of stored representations, and some object representation (of an individual or category) is activated. This explanatory picture rests on the Transformational Thesis: it amounts to matching one representation that makes explicit some information available in the retinal image with a stored representation, which by being activated, now makes explicit object-related information (e.g., category membership). So the very process of recognition appears to be one of reformatting.

Compare here an early criticism of Tarr and Bülthoff (1998, p.3) of object-centered theories and models of recognition:

> Reconstruction [of 3D shape from the 2D retinal image] assumes that visual perception is a hierarchical process which begins with local features that are combined into progressively more complex descriptions. . . Note that the types of features used and how they are combined is completely deterministic. That is, particular types of features and the relations between them are pre-defined and used for reconstruction across all images. Moreover, the presence or absence of a given feature is absolute—there is no 'middle ground' in which there is partial or probabilistic evidence for a feature.

However the problem is not that the approach makes recognition a deterministic process. Rather, it is that this picture is one of reformatting and hence making explicit, natural information latent in the retinal image. This is a problem, even if recognition allows for "partial or probabilistic" matches. In this respect, view-centered approaches fair no better. For example, DiCarlo and Cox (2007, p.334) state that:

> . . . one can view [object recognition] as the problem of finding operations that progressively transform [the] retinal representation into a new form of representation. . . ...

They go on to describe a view-centered approach to object recognition that characterizes object representations as manifolds in a high dimensional activation space. The idea is that manifolds for objects in the activation space of early visual areas are "entangled", and that the process of recognition involves constructing manifolds for objects that are linearly separable in activation space. When we construct a view-based representation of an object, it corresponds to a point in activation space, and based on position, is mapped to some object manifold. Whatever the merits of this as an approach to explaining the viewpoint invariance of object recognition (for a technical critique, see Goris and de Beeck, 2009), it overtly relies on the Transformational Thesis. Here is how DiCarlo and Cox (2007, p.335) describe the background of their view, with respect to a hypothetical identification task that requires discriminating between faces of two individuals:

> . . . although the retinal representation cannot directly support recognition, it implicitly contains the information to distinguish which of the two individuals was seen. We argue that this describes the computational crux of 'everyday' recognition: the problem is typically not a lack of information or noisy information, but that the information is badly formatted in the retinal representation—it is tangled. . .

So regardless of one's preferred approach (i.e., object-centered or view-centered), the process of recognition is characterized as a transformation and reformatting of natural information available in the retinal image. I believe that this shows that the disjunction problem cuts quite deep, as it reveals that the very structure of explanations of object recognition (regardless of the details) render them inadequate. And to some extent, this is as it should be. Recall that what was supposed to make recognition *hard* as a computational problem is that object representations manage to be viewpoint invariant and object specific given that the visual inputs are *both* underdetermining and variable with respect to their causes (Riesenhuber and Poggio, 2000; Rust and Stocker, 2010). One way of thinking about the upshot of the underdetermination problem is precisely that the information available from the retina is, at any given moment, *ambiguous*. It cannot be that vision is merely the process of reformatting natural information, unless one is willing to jettison the very idea of underdetermination. In fact, to some extent, this is precisely what the Transformational Thesis rests on—a rejection of the underdetermination problem (which Gibson, notoriously did reject; Epstein, 1977).

6.3.3 Argument Summary

If explanations of object recognition are to do better, they must reject (or revise) the Transformational Thesis, which means acknowledging that a crude informational theory will not do. Some strategies for solving the disjunction problem are required. But before briefly reviewing some of the most familiar proposals, let me take a moment to recap. The core of my argument in this chapter can be spelled out as follows:

(R1) Typical explanations of the viewpoint invariance and object specificity of object representations are adequate iff they entail that the content of object representations is viewpoint invariant and object specific.

(R2) The content of object representations is viewpoint invariant and object specific only if the content is perceptually robust.

(R3) If the content of a representation is determined solely by a reliable, information carrying causal relation, then the content is not perceptually robust.

(R4) Typical explanations of the invariance and specificity of object representations entail that the content of object representations is determined solely by a reliable, information carrying causal relation.

(R5) Typical explanations of the invariance and specificity of object representations entail that the content of object representations is not perceptually robust. [From (R3)-(R4)]

(R6) Hence, typical explanations of the invariance and specificity of object representations are not explanatorily adequate. [From (R1), (R2) and (R5)]

Given (R6), it is easy to now see why any augmented informational theory that solves—or provides a strategy for solving—the disjunction problem is quite relevant to explanations of object recognition. This was Conclusion 3 of my project. Far from reflecting obscure philosophical ruminations, the disjunction problem is a fundamental challenge to the adequacy of explanations of object recognition. More generally, it is a challenge to any explanation within the information-processing framework that assumes the Distal Object Thesis, and must posit perceptual representations, but offers theories and models of visual processing in line with the Transformational Thesis. Now let us look at how proposed solutions to the problem square with research on object recognition, and the story of vision more generally.

6.4 Can the Usual (Suspect) Solutions Aid Explanations of Object Recognition?

In this section I briefly assess some of the solutions usually proposed for the disjunction problem. All these solutions have a similar structure. Since the challenge is to determine why only some information carrying causal relations are determining of content, each solution presents an attempt to specify two sorts of contexts in which a representation is reliably tokened: the content determining ones, and the wild ones. This division can be between actual, or counterfactual states of affairs. Also, it is common to take the content-determining contexts to be ones that are in some sense normal or optimal.

Typically, theories of content are evaluated based on both metaphysical and explanatory constraints. I am not much concerned with the metaphysical constraints that are usually emphasized, namely, that the representing relation appeal solely to non-mental, natural ingredients (Fodor, 1984). Unless the view is viciously circular, I believe a theory of content can still be of use if it satisfies the desired explanatory constraints (cf. Rey, 2002). Along these lines I will primarily evaluate the usual proposals with respect to whether they are consistent with relevant facts about object recognition and the story of vision.

The three solutions I discuss appeal to either (i) learning, (ii) teleological function, or (iii) counterfactual conditionals. My discussion is far from exhaustive, as there are other possible avenues for addressing the disjunction problem that I will not consider.[10] And for the proposals I do discuss, there is more that can be said both in favor and against each of them. However, my aim is not to show decisively which is superior, but rather to get a sense of how they might be paired with explanations of object recognition. Thus for each I briefly summarize the approach, how it purports to solve the disjunction, and one objection to the approach which suggests that it is in tension with some background facts regarding the story of vision.

6.4.1 Learning and the Underdetermination Problem

Dretske (1981) was aware of the fact that not all reliable, information carrying relations could be content-determining. To address this problem, he appealed to learning conditions, the idea being that misrepresentation only occurs after the period in which a representation is acquired by an information-processing system. Here is how Dretske describes the idea, where the content in question is that some signal, s, instantiates the property F:[11]

> In the learning situation special care is taken to see that incoming signals have an intensity, a strength, sufficient unto delivering the required piece of information *to* the learning subject. . . Such precautions are taken in the learning situation. . . in order to ensure that an internal structure is developed with. . . the information that s is F. . . But once we have mean-

[10] For example two-factor approaches that include both an informational as well as conceptual role component to their theories.

[11] As quoted by Fodor (1987, p.102-103).

> ing, once the subject has articulated a structure that is selectively sensitive to information about the F-ness of things, instances of this structure, tokens of this type, can be triggered by signals that *lack* the appropriate piece of information... We [thus] have a case of misrepresentation—a token of a structure with a false content. We have, in a word, meaning without truth. (Dretske, 1981, p.194-195 emphasis in original)

This solution in effect appeals to optimal input conditions during the learning period. Since the content of a representation is acquired during the learning period, and the only reliable case of a representation during this period is what it is about (i.e., that s is F), wild tokenings of the representation outside the learning period are not content-determining. So the content of the representation is not disjunctive with respect to the causes of wild tokenings.

We have already seen in the last chapter why learning is important to achieving viewpoint invariance, so it is not implausible that it might play some role in characterizing how content is determined by a reliable causal relation. However, the immediate objection to Dretske's learning-based solution is that the distinction between the learning period, and the later period in which wild tokenings of a representation occur, is not principled (Fodor, 1984; Loewer, 1987). This objection is all the more damning with respect to our interests since part of the upshot of the underdetermination problem, as characterized in Chapter 4, is that there is never a circumstance in which only the content is a possible cause of tokenings of the representation. So the proposal is inconsistent with this part of the story of vision.

There are perhaps ways of addressing this concern. For example, if the learning period is supervised, then *if* a token of the representation had been caused by something other than what it represents, the “Teacher” would correct the learner. Fodor (1984, p.242) discusses this move, and suggests it does not work because it makes the content of representation dependent on the intentions of the Teacher. Ostensibly when it comes to perceptual representations, the “Teacher” is the world, so what is important is that there is simply feedback from the environment (cf. Godfrey-Smith, 1989). Whatever the merits of this suggestion, it is clear that if a learning-based solution is to help discharge the disjunction problem, we would need a more sophisticated story about how to demarcate the learning period (and process) than Dretske (1981) provides (Loewer, 1987, p.300-301).[12]

6.4.2 Teleological Function and the Distal World

Perhaps the most popular approach to theories of content are those that make some sort of appeal to teleological function. There is a great deal of variation among such theories (Neander, 012b), which include the teleological and consumer-based approach to generic representation that I discussed in Chapter 2 (Millikan, 1984, 1989; Papineau, 1987). Here I will focus on informational theories that appeal to teleological function to address the disjunction problem.

The general structure of the proposal is that some internal state represents

[12] At the limit, Dretske’s account might work for one-shot, or near one-shot, learning (Biederman and Bar, 1999), since in such a case there would be no chance of wild tokenings during the learning period.

some state of affairs if there is a reliable causal relation between the two, and the state has been selected to have the function of being caused by the state of affairs. Here is how Dretske (1988, p.65-66 emphasis in original) expresses the idea:[13]

> . . . it is important to remember that not every indicator, not even those that occur in plans and animals, is a representation, it is essential that it be the indicator's *function*. . . to indicate what it indicates. The width of growth rings in trees growing in semi-arid regions is a sensitive rain gauge, an accurate indication of the amount of rainfall in the year corresponding to the ring. This does not mean, however, that these rings *represent* the amount of rainfall in each year. For that to be the case, it would be necessary that it be the function of these rings to indicate, by their width, the amount of rain in the year corresponding to each ring.

What sort of function is at play? Typically it is taken to be a teleological function. How exactly we are to understand teleological function is one of the issues with this sort of move. Or for that matter, how we are to get the function attached to the particular state via teleology, or whether the function is to indicate (Godfrey-Smith, 1992). However, assuming this can be cashed out, misrepresentation occurs when a representation is tokened, and carries information about, something other than what it has the function of being caused by. Thus, loosely, instances of misrepresentation qualify as a form of malfunction (Neander, 1995).

[13] Note, an "indicator" is essentially equivalent to what I have been calling a sensory register, and Ramsey (2007) calls a receptor.

For my part, I think the very idea of appealing to teleological function is completely wrong headed when it comes to thinking about the constitutive conditions for the representing relation (I think it at most relates to the function of representations, rather than the constitutive conditions for intentional content; Cummins, 1996; Fodor, 1990). However, here I will just focus on one difficulty for informational theories that appeal to teleological function, namely what I will refer to as the "distality problem". The problem is that merely tracking properties of proximal stimulation might serve perfectly fine for adaptive purposes. So it is not clear how we get past proximal stimulation, and represent the distal world. Proponents of a teleological solution are well aware of this problem. For example, recently Neander (2012) has defended a teleological approach to perceptual representation, and offers the following account of why the teleological function of an internal state is to track the distal property, and not proximal input. Her solution is introduced with respect to an example of whether a toad represents worm-like motion of a distal object, or proximal patterns of input to the retina.

> The pathways in the toad's brain were selected for responding to both the distal worm-like motion and the more proximal patterns of light that carry information about the distal worm-like motion to the toad. But there is an important asymmetry here. These pathways in the toad's visual system were selected for responding to the light by producing certain tectal firings because *by that means* they respond to the distal worm-like motion, and not vice versa. That is, they were not selected

> for responding to the distal worm-like motion by producing certain tectal firings because *by that means* they responded to the more proximal patterns of light. That just isn't how the means-end analysis pans out. I believe this solves the problem of distal content. (Neander, 2012, p.34-35 emphasis in original)

Except Neander's proposal does no such thing. As Epstein (1977, p.6) points out, we might readily explain the presence of an adaptation in vision by how it allows an organism to respond to some survival-related variable in the distal world, but that does not tell us *what* it perceives. The same point applies when talking about perceptual representation. We can readily acknowledge the asymmetry pointed out by Neander, but it does nothing to show how teleological function gets us to perceptual objectivity. Of course this is just one proposal, but the point is that even granting that adding a teleological requirement addresses the disjunction problem (which itself is controversial to say the least), unless such a picture also addresses the distality problem, it will be inconsistent with some basic elements of the story of vision.

In contrast to Neander, Dretske (1986) offers a more promising solution to the distality problem, which relies on generalization across stimulus inputs, like I discussed with respect to perceptual learning in the previous chapter. However, for Dretske teleological function is not the basis for the solution to the distality problem.

6.4.3 Asymmetric Dependence and the Distal World

The last strategy for solving the disjunction problem that I will discuss is Fodor's (1987; 1990) asymmetric dependence account. While the view has few adherents I nonetheless believe it has some promise as a theory of content that could be paired with explanations of object recognition.[14] The basic idea is straightforward. The reason my concept CAT is about cats, and not a disjunction of its other causes is that the causal relations between, say, raccoons or possums on dark nights, and the concept is *asymmetrically dependent* on the relationship between cats and CAT. On the one hand, if cats did not cause tokenings of the concept, neither would these other creatures, while if raccoons and possums did not cause tokenings of the concept, cats still would. This counterfactual dependence breaks the symmetry between (Inf) and (Mis) from earlier, as the tokenings of a representation covered by (Mis) are in fact asymmetrically dependent on the nomic generalization (Inf).

I think that Fodor's account at least provides the right *sort* of conditions to supplement a crude informational theory. In effect, the proposal describes a higher-order causal dependency between causal dependencies (if one assumes, at least in part, a counterfactual theory of causation). The asymmetry shows that one dependency relation is the cause of the other, but not vice versa. Another virtue is that it is the only proposal explicitly based on a robust interpretation of the disjunction problem. As promising as I think asymmetric dependence is, it does face a problem related to the story of vision. As we have emphasized, the relationship

[14] Rare sympathetic discussions of the view can be found in Margolis (1998) and Rey (2009).

between perceptual representations and the distal world is causally mediated by proximal inputs. The problem is that it is not clear how asymmetric dependence is supposed to work in cases of causal mediation. Thus, it also faces a distality problem (Godfrey-Smith, 1989).

Fodor (1990, p.118) is seemingly aware of the problem. If occurrences of A reliably cause proximal inputs B and these inputs cause tokenings of the internal state C, then the dependency between A and C is asymmetrically dependent on the one between B and C. In which case, we need to know why it is that C does not represent the proximal inputs B. Fodor's initial take is to say that since the dependence of $A - C$ on $B - C$ is not robust, it is not "semantically relevant". Perhaps, but then so much the worse for asymmetric dependence. But the important case is actually one where we have two mediated causal chains, like in my example of misrepresentation. Raccoons and possums on dark nights cause cat-ish visual inputs, and cat-ish inputs cause tokenings of CAT. Cats of course also cause cat-ish inputs, and hence tokenings of CAT. Fodor says it is a "homework problem" to formulate his account such that the former mediated dependency is asymmetrically dependent on the latter.

However it is not obvious how to complete Fodor's homework assignment. Suppose raccoons and possums stopped looking cat-ish, and so no longer caused tokenings of CAT. This is one part of the homework. However, as pointed out by Prinz (2011), one might worry that it is not clear that the other necessary counterfactual holds: if cats stopped looking cat-ish, and no longer caused tokenings of the concept, it is not obvious why it would follow that raccoons and possums would not still be mistaken for cats. Presumably, they would still look as cat-ish as ever. But if asymmetric dependence is to work for cases of causal mediation, which is what is necessary for perceptual representation of the distal world, then we need a story for why this dependency would also be broken. Fodor never provides such a story, though I think it is possible that one could be developed.

In conclusion, the usual suspect solutions to the disjunction problem, in one way or another, seem to be in tension with basic facts about visual perception that are crucial background to explanations of object recognition. Although my discussion has been rather brief, I believe it is still possible that these proposals (and others I have not discussed) could be developed in a manner that makes them both consistent with the story of vision, and suitable solutions to the sort of disjunction problem faced by explanations of object recognition.

6.5 Summary and Conclusion

The ultimate aim of my project, recall, had two parts: first, to show that The Distal Object Thesis and Transformational Thesis, two common themes in the information-processing narrative in vision science, are in tension with each other; and second, to show that informational theories of content from philosophy offer strategies for relieving this tension. This chapter served to meet these ultimate aims. The root of the tension, as we have seen, is the disjunction problem, and various strategies are available for addressing the problem, though how they might be married to explanations of object recognition remains to be seen.

Along the way to achieving these aims I defended the indispensability of a notion of internal representation to the information-processing framework in general. This involved a careful, and I think, useful, analysis of the influential work of Marr (1982). Also, to show why the Distal Object Thesis is essential to explanations of object vision, I showed how a notion of perceptual representation is indispensable to explanations of object recognition. This involved retrofitting the argument from constancy, a classic strategy for illustrating the objectivity of perception.

I think these arguments and conclusions are philosophically interesting in their own right. They show why a notion of perceptual (mental) representation is fundamental to some of the explanations in vision science, and why philosophical theories of content, which are somewhat out of fashion at present, might make a real contribution to addressing a problem that these explanations face. More generally, they show why questions of content are of central importance to how vision science explains what we see, and how we see it.

Bibliography

Adams, F. and Aizawa, K. (2010). Causal theories of mental content. In Zalta, E. N., editor, *Stanford Encyclopedia of Philosophy.* http://plato.stanford.edu/entries/content-causal/.

Afraz, A., Pashkam, M. V., and Cavanagh, P. (2010). Spatial heterogeneity in the perception of face and form attributes. *Current Biology*, 20(23):2112–2116.

Afraz, S.-R., Kiani, R., and Esteky, H. (2006). Microstimulation of inferotemporal cortex influences face categorization. *Nature*, 442(7103):692–695.

Akins, K. (1996). Of sensory systems and the" aboutness" of mental states. *The Journal of Philosophy*, 93(7):337–372.

Ashby, F. G. and Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, 93(2):154.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, 61(3):183.

Barenholtz, E. and Tarr, M. J. (2006). Reconsidering the role of structure in vision. *BH Ross (Series Ed.) & AB Markman (Vol. Ed.), Categories in use: The psychology of learning and motivation*, 47:157–180.

Barlow, H. . (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, 1:371–394.

Barlow, H. (2001). Redundancy reduction revisited. *Network: Computation in neural systems*, 12(3):241–253.

Barlow, H. B. (1953). Summation and inhibition in the frog's retina. *The Journal of Physiology*, 119(1):69–88.

Bechtel, W. (1998). Representations and cognitive explanations: Assessing the dynamicist's challenge in cognitive science. *Cognitive Science*, 22(3):295–318.

Bechtel, W. (2008). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. Taylor & Francis.

Bechtel, W. and Shagrir, O. (2015). The non-redundant contributions of Marr's three levels of analysis for explaining information-processing mechanisms. *Topics in Cognitive Science*, 7(2):312–322.

Bevan, W. (1958). Perception: Evolution of a concept. *Psychological Review*, 65(1):34.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review*, 94(2):115.

Biederman, I. (2000). Recognizing depth-rotated objects: A review of recent research and theory. *Spatial Vision*, 13(2):241–253.

Biederman, I. and Bar, M. (1999). One-shot viewpoint invariance in matching novel objects. *Vision Research*, 39(17):2885–2899.

Biederman, I. and Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human perception and performance*, 19(6):1162.

Biederman, I., Subramaniam, S., Bar, M., Kalocsai, P., and Fiser, J. (1999). Subordinate-level object classification reexamined. *Psychological Research*, 62(2-3):131–153.

Blanz, V., Tarr, M. J., Bülthoff, H. H., and Vetter, T. (1999). What object attributes determine canonical views? *Perception-London*, 28(5):575–600.

Blaser, E., Pylyshyn, Z. W., and Holcombe, A. O. (2000). Tracking an object through feature space. *Nature*, 408(6809):196–199.

Block, N. (1986). Advertisement for a semantics for psychology. *Midwest Studies in Philosophy*, 10(1):615–78.

Boring, E. G. (1946). Perception of objects. *American Journal of Physics*, 14:99–107.

Boring, E. G. (1952). Visual perception as invariance. *Psychological Review*, 59(2):141.

Brentano, F. (1874). *Psychologie vom empirischen Standpunkte*, volume 1. Duncker & Humblot.

Brunswik, E. (1940). Thing constancy as measured by correlation coefficients. *Psychological Review*, 47(1):69.

Bülthoff, H. H. and Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences*, 89(1):60–64.

Burge, T. (1986). Individualism and psychology. *The Philosophical Review*, 95(1):3–45.

Burge, T. (2003). Perceptual entitlement*. *Philosophy and Phenomenological Research*, 67(3):503–548.

Burge, T. (2010). *Origins of objectivity*. Oxford University Press, Oxford, UK.

Cadieu, C. F., Hong, H., Yamins, D. L., Pinto, N., Ardila, D., Solomon, E. A., Majaj, N. J., and DiCarlo, J. J. (2014). Deep neural networks rival the representation of primate it cortex for core visual object recognition. *PLoS Computational Biology*, 10(12):e1003963.

Canny, J. (1986). A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 8(6):679–698.

Cao, R. (2012). A teleosemantic approach to information in the brain. *Biology & Philosophy*, 27(1):49–71.

Casati, R. (2015). Object perception. In Matthen, M., editor, *The Oxford Handbook of the Philosophy of Perception*, pages 393–404. Oxford University Press, USA.

Cassirer, E. (1944). The concept of group and the theory of perception. *Philosophy and Phenomenological Research*, 5(1):1–36.

Cheung, O. S., Hayward, W. G., and Gauthier, I. (2009). Dissociating the effects of angular disparity and image similarity in mental rotation and object recognition. *Cognition*, 113(1):128–133.

Chisholm, R. M. (1952). Intentionality and the theory of signs. *Philosophical Studies*, 3(June):56–63.

Chomsky, N. (1995). Language and nature. *Mind*, 104(413):1–61.

Chomsky, N. (2000). *New horizons in the study of language and mind.* Cambridge University Press, Cambridge, UK.

Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *The Journal of Philosophy*, 78(2):67–90.

Cohen, J. (2004). Information and content. In Floridi, L., editor, *The Blackwell guide to the philosophy of computing and information*, pages 215–227. Blackwell Publishing Ltd.

Cohen, J. (2008). Colour constancy as counterfactual. *Australasian Journal of Philosophy*, 86(1):61–92.

Cohen, J. (2015). Perceptual representation, veridicality, and the interface theory of perception. *Psychonomic Bulletin & Review*, online first(1–7).

Cohen, J. and Meskin, A. (2006). An objective counterfactual theory of information. *Australasian Journal of Philosophy*, 84(3):333–352.

Coltheart, V., Mondy, S., and Coltheart, M. (2005). Repetition blindness for novel objects. *Visual Cognition*, 12(3):519–540.

Cooper, L. A. and Hochberg, J. (1994). Objects of the mind: Mental representations in visual perception and cognition. In Ballesteros, S., editor, *Cognitive approaches to human perception*. Erlbaum Hillsdale, NJ.

Coren, S. and Girgus, J. S. (1977). Illusions and constancies. In Epstein, W., editor, *Stability and constancy in visual perception: Mechanisms and processes*, pages 255–283. Wiley & Sons.

Cox, D. D. (2014). Do we understand high-level vision? *Current Opinion in Neurobiology*, 25:187–193.

Cox, D. D., Meier, P., Oertelt, N., and DiCarlo, J. J. (2005). 'Breaking' position-invariant object recognition. *Nature Neuroscience*, 8(9):1145–1147.

Crane, T. (1998). Intentionality as the mark of the mental. *Royal Institute of Philosophy Supplement*, 43:229–251.

Craver, C. F. (2001). Role functions, mechanisms, and hierarchy. *Philosophy of Science*, 68(1):53–74.

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

Cummins, R. (1989). *Meaning and mental representation.* MIT Press, Cambridge, MA.

Cummins, R. (1996). *Representations, targets, and attitudes.* MIT press.

Cummins, R. C. (1975). Functional analysis. *Journal of Philosophy*, 72:741–764.

Cummins, R. C. (1983). *The nature of psychological explanation.* MIT Press,.

Cutting, J. E. (1982). Two ecological perspectives: Gibson vs. Shaw and Turvey. *The American Journal of Psychology*, 95:199–222.

Cutting, J. E. (1983). Four assumptions about invariance in perception. *Journal of Experimental Psychology: Human Perception and Performance*, 9(2):310.

Cutzu, F. and Edelman, S. (1994). Canonical views in object representation and recognition. *Vision Research*, 34(22):3037–3056.

Darden, L. (1991). *Theory Change in Science: Strategies From Mendelian Genetics.* Oxford UniveMIT Press, New York.

Davies, M. (1991). Individualism and perceptual content. *Mind*, 100(399):461–84.

Dewey, J. (1896). Psychology of number. *Science*, pages 286–289.

DiCarlo, J. J. and Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences*, 11(8):333–341.

DiCarlo, J. J., Zoccolan, D., and Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, 73(3):415–434.

Dretske, F. (1981). *Knowledge and the Flow of Information.* The MIT Press, Cambridge, MA.

Dretske, F. (1986). Misrepresentation. In Bogdan, R., editor, *Belief: Form, Content, and Function*, pages 17–36. Oxford University Press.

Dretske, F. I. (1983). Precis of knowledge and the flow of information. *Behavioral and Brain Sciences*, 6(1):55–63.

Dretske, F. I. (1988). *Explaining behavior: Reasons in a world of causes.* MIT Press, Cambridge, MA.

Edelman, S. and Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research*, 32(12):2385–2400.

Egan, F. (1992). Individualism, computation, and perceptual content. *Mind*, 101(403):443–59.

Egan, F. (1995). Computation and content. *The Philosophical Review*, 104(2):181–203.

Egan, F. (1999). In defence of narrow mindedness. *Mind and Language*, 14(2):177–94.

Egan, F. (2010). Computational models: a modest role for content. *Studies in History and Philosophy of Science Part A*, 41(3):253–259.

Elliott, K. (2004). Error as means to discovery. *Philosophy of Science*, 71(2):174–197.

Enç, B. (1982). Intentional states of mechanical devices. *Mind*, 91(362):161–183.

Epstein, W. (1973). The process of taking-into-account? in visual perception. *Perception*, 2(3):267–285.

Epstein, W. (1977). Historical introduction to the constancies. In Epstein, W., editor, *Stability and constancy in visual perception: Mechanisms and processes*, pages 1–22. Wiley & Sons.

Epstein, W., Park, J., and Casey, A. (1961). The current status of the size-distance hypotheses. *Psychological Bulletin*, 58(6):491–514.

Evans, G. (1982). *Varieties of Reference.* Oxford University Press,.

Feldman, J. (2003). What is a visual object? *Trends in Cognitive Sciences*, 7(6):252–256.

Feldman, J. (2012). Symbolic representation of probabilistic worlds. *Cognition*, 123(1):61–83.

Feldman, J. and Tremoulet, P. D. (2006). Individuation of visual objects over time. *Cognition*, 99(2):131–165.

Field, H. H. (1978). Mental representation. *Erkenntnis*, 13(1):9–61.

Flombaum, J. I., Scholl, B. J., and Santos, L. R. (2009). Spatiotemporal priority as a fundamental principle of object persistence. In Hood, B. M. and Santos, L., editors, *The Origins of Object Knowledge*, pages 135–164. Oxford University Press, New York.

Fodor, J. (2000). *The mind does not work that way.* MIT Press, Cambridge, MA.

Fodor, J. A. (1968). *Psychological Explanation: An Introduction To The Philosophy Of Psychology.* Random House, New York.

Fodor, J. A. (1975). *The language of thought.* Harvard University Press, Cambridge, MA.

Fodor, J. A. (1980). Methodological solipsism considered as a research strategy in cognitive psychology. *Behavioral and Brain Sciences*, 3(1):63–73.

Fodor, J. A. (1984). Semantics, Wisconsin style. *Synthese*, 59(3):231–250.

Fodor, J. A. (1986). Why paramecia don't have mental representations. *Midwest Studies in Philosophy*, 10(1):3–23.

Fodor, J. A. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind.* The MIT Press.

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

Fodor, J. A. (1990b). Information and representation. In Hanson, P. P., editor, *Information, Language and Cognition*. UniOxford University Press, New York, Vancouver.

Foster, D. H. (2003). Does colour constancy exist? *Trends in Cognitive Sciences*, 7(10):439–443.

Gallistel, C. R. and King, A. P. (2009). *Memory and the computational brain: Why cognitive science will transform neuroscience*, volume 6. Wiley-Blackwell, Malden, MA.

Ganson, T., Bronner, B., and Kerr, A. (2014). Burge's defense of perceptual content. *Philosophy and Phenomenological Research*, 88(3):556–573.

Gauthier, I., Hayward, W. G., Tarr, M. J., Anderson, A. W., Skudlarski, P., and Gore, J. C. (2002). Bold activity during mental rotation and viewpoint-dependent object recognition. *Neuron*, 34(1):161–171.

Gibson, J. J. (1950). *The Perception Of The Visual World*. Houghton Mifflin, Boston, MA.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston, MA.

Gillam, B. (1998). Illusions at century's end. In Hochberg, J., editor, *Perception and cognition at century's end*, pages 95–136. Academic Press, San Diego.

Godfrey-Smith, P. (1989). Misinformation. *Canadian Journal of Philosophy*, 19(4):533–50.

Godfrey-Smith, P. (1991). Signal, decision, action. *The Journal of Philosophy*, 88(12):709–722.

Godfrey-Smith, P. (1992). Indication and adaptation. *Synthese*, 92(2):283–312.

Goris, R. L. and de Beeck, H. P. O. (2009). Neural representations that support invariant object recognition. *Frontiers in Computational Neuroscience*, 3.

Green, D. M. and Swets, J. A. (1966). *Signal detection theory and psychophysics.* John Wiley, New York.

Grice, H. P. (1957). Meaning. *Philosophical Review*, 66(3):377–388.

Griffiths, P. E. (1997). *What emotions really are: The problem of psychological categories.* Chicago University Press, Chicago.

Griffiths, P. E. (2004). Emotions as natural and normative kinds. *Philosophy of Science*, 71(5):901–911.

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., and Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences*, 14(8):357–364.

Gwynne, D. and Rentz, D. (1983). Beetles on the bottle: male buprestids mistake stubbies for females (coleoptera). *Australian Journal of Entomology*, 22(1):79–80.

Harman, G. (1982). Conceptual role semantics. *Notre Dame Journal of Formal Logic*, 23(2):242–256.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1):335–346.

Hatfield, G. (2002). Perception as unconscious inference. In Heyer, D., editor, *Perception and the Physical World: Psychological and Philosophical Issues in Perception*, pages 113–143. John Wiley and Sons, New York.

Haugeland, J. (1991). Representational genera. In Stich, W. R. S. and Rumelhart, D., editors, *Philosophy and connectionist theory*, pages 61–89. Erlbaum Hillsdale, NJ.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539):2425–2430.

Hayward, W. G. (2003). After the viewpoint debate: where next in object recognition? *Trends in Cognitive Sciences*, 7(10):425–427.

Hayward, W. G. (2012). Whatever happened to object-centered representations? *Perception*, 41(9):1153.

Hayward, W. G. and Tarr, M. J. (1997). Testing conditions for viewpoint invariance in object recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 23(5):1511.

Hayward, W. G. and Williams, P. (2000). Viewpoint dependence and object discriminability. *Psychological Science*, 11(1):7–12.

Heck, R. (2000). Nonconceptual content and the "space of reasons". *Philosophical Review*, 109(4):483–523.

Hein, E. and Moore, C. M. (2012). Spatio-temporal priority revisited: The role of feature identity and similarity for object correspondence in apparent motion. *Journal of Experimental Psychology: Human Perception and Performance*, 38(4):975.

Hilbert, D. and Byrne, A. (2003). Color perception and color science. *Behavioral and Brain Sciences*, 26(1):3–21.

Hochberg, J. (1998). *Perception and Cognition at Century's End: History, Philosophy, Theory.* Academic Press.

Hochberg, J. E. (1957). Effects of the gestalt revolution: The cornell symposium on perception. *Psychological Review*, 64(2):73.

Hoffman, D. D. and Richards, W. A. (1984). Parts of recognition. *Cognition*, 18(1):65–96.

Hoffman, D. D. and Singh, M. (2012). Computational evolutionary perception. *Perception*, 41(9):1073.

Hollingworth, A. and Franconeri, S. L. (2009). Object correspondence across brief occlusion is established on the basis of both spatiotemporal and surface feature cues. *Cognition*, 113(2):150–166.

Holway, A. H. and Boring, E. G. (1941). Determinants of apparent visual size with distance variant. *The American Journal of Psychology*, pages 21–37.

Ittelson, W. H. (1951). The constancies in perceptual theory. *Psychological Review*, 58(4):285.

Kahneman, D., Treisman, A., and Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, 24(2):175–219.

Keeley, B. L. (2002). Making sense of the senses: Individuating modalities in humans and other animals. *Journal of Philosophy*, 99(1):5–28.

Kersten, D. and Yuille, A. (2003). Bayesian models of object perception. *Current Opinion in Neurobiology*, 13(2):150–158.

Kilpatrick, F. and Ittelson, W. (1953). The size-distance invariance hypothesis. *Psychological Review*, 60(4):223.

Kilpatrick, F. P. (1954). Two processes in perceptual learning. *Journal of Experimental Psychology*, 47(5):362.

Kirchner, H. and Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research*, 46(11):1762–1776.

Kirsh, D. (1992). When is information explicitly represented? *The Vancouver Studies in Cognitive Science*, pages 340–365.

Koffka, K. (1935). *Principles of Gestalt psychology.* Harcourt, Brace.

Kravitz, D. J., Vinson, L. D., and Baker, C. I. (2008). How position dependent is visual object recognition? *Trends in Cognitive Sciences*, 12(3):114–122.

Kriegeskorte, N. and Kievit, R. A. (2013). Representational geometry: integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, 17(8):401–412.

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., and Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, 60(6):1126–1141.

Kripke, S. A. (1980). *Naming and Necessity.* Harvard University Press.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

Leslie, A. M., Xu, F., Tremoulet, P. D., and Scholl, B. J. (1998). Indexing and the object concept: developingwhat'andwhere'systems. *Trends in Cognitive Sciences*, 2(1):10–18.

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

Liu, Z., Knill, D. C., and Kersten, D. (1995). Object classification for human and ideal observers. *Vision Research*, 35(4):549–568.

Loewer, B. (1987). From information to intentionality. *Synthese*, 70(2):287–317.

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

Machery, E. (2005). Concepts are not a natural kind*. *Philosophy of Science*, 72(3):444–467.

Machery, E. et al. (2009). *Doing without concepts.* Oxford University Press, New York.

Maley, C. J. (2011). Analog and digital, continuous and discrete. *Philosophical Studies*, 155(1):117–131.

Mallon, R., Machery, E., Nichols, S., and Stich, S. (2009). Against arguments from reference. *Philosophy and Phenomenological Research*, 79(2):332–356.

Margolis, E. (1998). How to acquire a concept. *Mind and Language*, 13(3):347–369.

Mark, J. T., Marion, B. B., and Hoffman, D. D. (2010). Natural selection and veridical perceptions. *Journal of Theoretical Biology*, 266(4):504–515.

Marr, D. (1976). Early processing of visual information. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 275(942):483–519.

Marr, D. (1982). *Vision.* Freeman and Company, San Francisco.

Marr, D. and Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London B: Biological Sciences*, 207(1167):187–217.

Marr, D. and Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London B: Biological Sciences*, 200(1140):269–294.

Matthen, M. (2015). The individuation of the senses. In Matthen, M., editor, *Oxford Handbook of the Philosophy of Perception*, pages 567–586. Oxford University Press.

Matthen, M. P. (2005). *Seeing, Doing, and Knowing: A Philosophical Theory of Sense Perception*. Oxford University Press.

Mendelovici, A. (2013). Reliable misrepresentation and tracking theories of mental representation. *Philosophical Studies*, 165(2):421–443.

Millikan, R. G. (1984). *Language, thought, and other biological categories: New foundations for realism*. MIT press.

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

Mitroff, S. R. and Alvarez, G. A. (2007). Space and time, not surface features, guide object persistence. *Psychonomic Bulletin & Review*, 14(6):1199–1204.

Moore, C. M., Stephens, T., and Hein, E. (2010). Features, as well as space and time, guide object persistence. *Psychonomic Bulletin & Review*, 17(5):731–736.

Morgan, A. (2014). Representations gone mental. *Synthese*, 191(2):213–244.

Mur, M., Bandettini, P. A., and Kriegeskorte, N. (2009). Revealing representational content with pattern-information fmrian introductory guide. *Social, Cognitive, and Affective Neuroscience*, pages 1–9.

Nagel, T. (1980). The limits of objectivity. *The Tanner Lectures on Human Values*, 1:75–139.

Nanay, B. (2015). Perceptual representation / perceptual content. In Matthen, M., editor, *Oxford Handbook for the Philosophy of Perception*, pages 153–167. Oxford University Press.

Neander, K. (2012b). Teleological theories of mental content. In Zalta, E. N., editor, *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/content-teleological/.

Neander, K. L. (1995). Misrepresenting & malfunctioning. *Philosophical Studies*, 79(2):109–141.

Neander, K. L. (2012). Toward an informational teleosemantics. In Klingsbury, J. and D.Ryder, editors, *Millikan and Her Critics*. Blackwell.

Newell, A. (1980). Physical symbol systems. *Cognitive Science*, 4(2):135–183.

Newell, A. (1982). The knowledge level. *Artificial Intelligence*, 18(1):87–127.

Nguyen, A., Yosinski, J., and Clune, J. (2014). Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. *arXiv preprint arXiv:1412.1897*.

Norman, K. A., Polyn, S. M., Detre, G. J., and Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fmri data. *Trends in Cognitive Sciences*, 10(9):424–430.

Olin, L. (2014). Burge on perception and sensation. *Synthese*, pages 1–30.

O'Regan, J. K. and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(5):883–917.

Orlandi, N. (2014). *The innocent eye: why vision is not a cognitive process.* Oxford University Press, New York.

Palmer, S. (1978). Fundamental aspects of cognitive representation. In Rosch, E. and Lloyd, B., editors, *Cognition and Categorization*, pages 259–303. Lawrence Elbaum Associates.

Palmer, S. E. (1999). *Vision science: Photons to phenomenology.* MIT Press Cambridge, MA.

Papineau, D. (1987). *Reality and representation.* Blackwell, Oxford, UK.

Peacocke, C. (1994). Content, computation and externalism. *Mind & Language*, 9(3):303–335.

Peissig, J. J. and Tarr, M. J. (2007). Visual object recognition: do we know more now than we did 20 years ago? *Annual Review of Psychology*, 58:75–96.

Pereira, F., Mitchell, T., and Botvinick, M. (2009). Machine learning classifiers and fmri: a tutorial overview. *Neuroimage*, 45(1):S199–S209.

Peterson, M. A., Gillam, B., and Sedgwick, H. (2006). *In the mind's eye: Julian Hochberg on the perception of pictures, films, and the world*. Oxford University Press.

Piccinini, G. (2007). Computing mechanisms. *Philosophy of Science*, 74(4):501–526.

Piccinini, G. (2008). Computation without representation. *Philosophical Studies*, 137(2):205–241.

Piccinini, G. (2010). The mind as neural software? understanding functionalism, computationalism, and computational functionalism. *Philosophy and Phenomenological Research*, 81(2):269–311.

Piccinini, G. and Craver, C. (2011). Integrating psychology and neuroscience: Functional analyses as mechanism sketches. *Synthese*, 183(3):283–311.

Piccinini, G. and Scarantino, A. (2010). Computation vs. information processing: why their difference matters to cognitive science. *Studies in History and Philosophy of Science Part A*, 41(3):237–246.

Picciuto, V. and Carruthers, P. (2013). Inner sense. In S., B., M., M., and D., S., editors, *Perception and its Modalites*. Oxford University Press.

Prinz, J. (2011). Has mentalese earned its keep? on Jerry Fodor's lot 2. *Mind*, 120(478):485–501.

Pylyshyn, Z. (1989). The role of location indexes in spatial perception: A sketch of the finst spatial-index model. *Cognition*, 32(1):65–97.

Pylyshyn, Z. (2004). Some puzzling findings in multiple object tracking: I. tracking without keeping track of object identities. *Visual Cognition*, 11(7):801–822.

Pylyshyn, Z. W. (1984). *Computation and cognition.* Cambridge Univ Press.

Pylyshyn, Z. W. (2007). *Things and Places: How the Mind Connects with the World.* The MIT Press.

Pylyshyn, Z. W. and Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, 3(3):179–197.

Quine, W. V. (1960). *Word and object.* MIT Press, Cambridge, MA.

Ramsey, W. M. (2007). *Representation reconsidered.* Cambridge University Press, New York.

Rescorla, M. (2014). The causal relevance of content to computation. *Philosophy and Phenomenological Research*, 88(1):173–208.

Rey, G. (1997). *Contemporary philosophy of mind: A contentiously classical approach.* Blackwell,.

Rey, G. (1998). Keeping meaning in mind. In Pylyshyn, Z., editor, *Constraining Cognitive Theories: Issues and Options*. Greenwood Publishing Group.

Rey, G. (2002). Physicalism and psychology: A plea for substantive philosophy of mind. In Gillet, G. and Loewer, B., editors, *Physicalism and its discontents.* Cambridge University Press, Cambridge.

Rey, G. (2009). Concepts, defaults, and internal asymmetric dependencies: Distillations of fodor and horwich. In Suhm, N. K. C. C., editor, *The A Priori and its role in philosophy*, pages 185–204. Paderborn: Mentis.

Rey, G. (2012). Externalism and inexistence in early content. In Schantz, R., editor, *Prospects for Meaning*. De Gruyter.

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

Riesenhuber, M. and Poggio, T. (2000). Models of object recognition. *Nature Neuroscience*, 3:1199–1204.

Ritchie, J., Klein, C., and Kaplan, D. (unpublished). Internal representation and the decoder's dictum in cognitive neuroscience. Macquarie University.

Rock, I. (1983). *The Logic Of Perception*. MIT Press, Cambridge, MA.

Ross, H. E. and Plug, C. (1998). The history of size constancy and size illusions. In Walsh, V. and Kulikowski, J., editors, *Perceptual Constancy*, pages 499–528. Cambridge University Press.

Rupert, R. D. (2008). Causal theories of mental content. *Philosophy Compass*, 3(2):353–380.

Rust, N. C. and DiCarlo, J. J. (2010). Selectivity and tolerance (invariance?) both increase as visual information propagates from cortical area v4 to it. *The Journal of Neuroscience*, 30(39):12978–12995.

Rust, N. C. and Stocker, A. A. (2010). Ambiguity and invariance: two fundamental challenges for visual processing. *Current Opinion in Neurobiology*, 20(3):382–388.

Scarantino, A. (2015). Information as a probabilistic difference maker. *Australasian Journal of Philosophy*, 93(3):419–443.

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, 80(1):1–46.

Scholl, B. J. (2009). What have we learned about attention from multiple object tracking (and vice versa). In Dedrick, D. and Trick, L., editors, *Computation, cognition, and Pylyshyn*, pages 49–78. MIT Press, Cambridge, MA.

Scholl, B. J., Pylyshyn, Z. W., and Feldman, J. (2001). What is a visual object? evidence from target merging in multiple object tracking. *Cognition*, 80(1):159–177.

Schurgin, M. W., Reagh, Z. M., Yassa, M. A., and Flombaum, J. I. (2013). Spatiotemporal continuity alters long-term memory representation of objects. *Visual Cognition*, 21(6):715–718.

Schyns, P. G. (1998). Diagnostic recognition: task constraints, object information, and their interactions. *Cognition*, 67(1):147–179.

Segal, G. (1989). Seeing what is not there. *The Philosophical Review*, 98(2):189–214.

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., and Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(3):411–426.

Shagrir, O. (2010). Marr on computational-level theories. *Philosophy of Science*, 77(4):477–500.

Shagrir, O. (2012). Structural representations and the brain. *British Journal for the Philosophy of Science*, 63(3):519–545.

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423.

Shapiro, L. A. (1997). A clearer vision. *Philosophy of Science*, 64(1):131–53.

Shea, N. (2013). Naturalising representational content. *Philosophy Compass*, 8(5):496–509.

Shepard, R. N. (1984). Ecological constraints on internal representation: resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological Review*, 91(4):417.

Skyrms, B. (2010). *Signals: Evolution, Learning, and Information*. Oxford University Press, Oxford, UK.

Sprevak, M. (2011). Review william m. ramsey *Representation Reconsidered*. *The British Journal for the Philosophy of Science*, pages 1–7.

Stampe, D. W. (1977). Toward a causal theory of linguistic representation1. *Midwest Studies in Philosophy*, 2(1):42–63.

Stankiewicz, B. J. (2002). Empirical evidence for independent dimensions in the visual representation of three-dimensional shape. *Journal of Experimental Psychology: Human Perception and Performance*, 28(4):913.

Stankiewicz, B. J. (2003). Just another view. *Trends in Cognitive Sciences*, 7(12):526–526.

Sterelny, K. (1990). *The representational theory of mind.* Blackwell, Oxford, UK.

Stevens, S. S. (1951). Mathematics, measurement, and psychophysics. In Stevens, S., editor, *Handbook of Experimental Psychology.* Wiley and Sons, Oxford.

Stich, S. (1992). What is a theory of mental representation? *Mind*, pages 243–261.

Stich, S. P. (1983). *From folk psychology to cognitive science: The case against belief.* MIT Press, Cambridge, MA.

Stich, S. P. (1996). *Deconstructing the mind.* Oxford University Press, New York.

Stojanoski, B. and Cusack, R. (2014). Time to wave good-bye to phase scrambling: Creating controlled scrambled images using diffeomorphic transformations. *Journal of Vision*, 14(12):6.

Swoyer, C. (1991). Structural representation and surrogative reasoning. *Synthese*, 87(3):449–508.

Taddeo, M. and Floridi, L. (2005). Solving the symbol grounding problem: a critical review of fifteen years of research. *Journal of Experimental & Theoretical Artificial Intelligence*, 17(4):419–445.

Tarr, M. J. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin & Review*, 2(1):55–82.

Tarr, M. J. and Bülthoff, H. H. (1995). Is human object recognition better described by geon structural descriptions or by multiple views? comment on biederman and gerhardstein (1993). *Journal of Experimental Psychology: Human Perception and Performance*, 21(6):1494–1505.

Tarr, M. J. and Bülthoff, H. H. (1998). Image-based object recognition in man, monkey and machine. *Cognition*, 67(1):1–20.

Tarr, M. J., Bülthoff, H. H., Zabinski, M., and Blanz, V. (1997). To what extent do unique parts influence recognition across changes in viewpoint? *Psychological Science*, 8(4):282–289.

Tarr, M. J. and Gauthier, I. (1998). Do viewpoint-dependent mechanisms generalize across members of a class? *Cognition*, 67(1):73–110.

Tarr, M. J., Kersten, D., and Bülthoff, H. H. (1998a). Why the visual recognition system might encode the effects of illumination. *Vision Research*, 38(15):2259–2275.

Tarr, M. J. and Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, 21(2):233–282.

Tarr, M. J., Williams, P., Hayward, W. G., and Gauthier, I. (1998b). Three-dimensional object recognition is viewpoint dependent. *Nature Neuroscience*, 1(4):275–277.

Ternus, J. (1926). Experimentelle untersuchungen über phänomenale identität. *Psychologische Forschung*, 7(1):81–136.

Thorpe, S., Fize, D., Marlot, C., et al. (1996). Speed of processing in the human visual system. *Nature*, 381(6582):520–522.

Thouless, R. H. (1931). Phenomenal regression to the real object. i. *British Journal of Psychology. General Section*, 21(4):339–359.

Tian, M. and Grill-Spector, K. (2015). Spatiotemporal information during unsupervised learning enhances viewpoint invariant object recognition. *Journal of Vision*, 15(6).

Treisman, A. M. and Kanwisher, N. G. (1998). Perceiving visually presented objets: recognition, awareness, and modularity. *Current Opinion in Neurobiology*, 8(2):218–226.

Turvey, M. T., Shaw, R. E., Reed, E. S., and Mace, W. M. (1981). Ecological laws of perceiving and acting: In reply to fodor and pylyshyn (1981). *Cognition*, 9(3):237–304.

Tye, M. (1992). Naturalism and the mental. *Mind*, 101(403):421–441.

Tye, M. (2000). *Consciousness, Color, and Content.* MIT Press, Cambridge, MA.

Ullman, S. (1979). *The interpretation of visual motion.* MIT Press, Cambridge, MA.

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

Vuong, Q. C. and Tarr, M. J. (2004). Rotation direction affects object recognition. *Vision Research*, 44(14):1717–1730.

Wallis, G., Backus, B. T., Langer, M., Huebner, G., and Bülthoff, H. (2009). Learning illumination-and orientation-invariant representations of objects through temporal association. *Journal of vision*, 9(7):6.

Wallis, G. and Bülthoff, H. H. (2001). Effects of temporal association on recognition memory. *Proceedings of the National Academy of Sciences*, 98(8):4800–4804.

Walsh, V. and Kulikowski, J. (1998). *Perceptual constancy: Why things look as they do.* Cambridge Univ Press, Cambridge, UK.

Withagen, R. and Chemero, A. (2009). Naturalizing perception developing the gibsonian approach to perception along evolutionary lines. *Theory & Psychology*, 19(3):363–389.

Yilmaz, H. (1967). Perceptual invariance and the psychophysical law. *Perception & Psychophysics*, 2(11):533–538.

Zacks, J. M. (2008). Neuroimaging studies of mental rotation: a meta-analysis and review. *Journal of Cognitive Neuroscience*, 20(1):1–19.