

REFERENCE ONLY

UNIVERSITY OF LONDON THESIS

Degree PhD

Year 2006

Name of Author HONSEY, A. S.

**COPYRIGHT**

This is a thesis accepted for a Higher Degree of the University of London. It is an unpublished typescript and the copyright is held by the author. All persons consulting the thesis must read and abide by the Copyright Declaration below.

**COPYRIGHT DECLARATION**

I recognise that the copyright of the above-described thesis rests with the author and that no quotation from it or information derived from it may be published without the prior written consent of the author.

**LOANS**

Theses may not be lent to individuals, but the Senate House Library may lend a copy to approved libraries within the United Kingdom, for consultation solely on the premises of those libraries. Application should be made to: Inter-Library Loans, Senate House Library, Senate House, Malet Street, London WC1E 7HU.

**REPRODUCTION**

University of London theses may not be reproduced without explicit written permission from the Senate House Library. Enquiries should be addressed to the Theses Section of the Library. Regulations concerning reproduction vary according to the date of acceptance of the thesis and are listed below as guidelines.

- A. Before 1962. Permission granted only upon the prior written consent of the author. (The Senate House Library will provide addresses where possible).
- B. 1962 - 1974. In many cases the author has agreed to permit copying upon completion of a Copyright Declaration.
- C. 1975 - 1988. Most theses may be copied upon completion of a Copyright Declaration.
- D. 1989 onwards. Most theses may be copied.

***This thesis comes within category D.***

This copy has been deposited in the Library of VCL

This copy has been deposited in the Senate House Library, Senate House, Malet Street, London WC1E 7HU.



# **The Content and Acquisition of Lexical Concepts**

*Richard Samuel Horsey*

Thesis submitted in partial fulfilment of the requirements  
for the degree of Doctor of Philosophy

University College London, January 2006

UMI Number: U592909

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U592909

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.  
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against  
unauthorized copying under Title 17, United States Code.



ProQuest LLC  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106-1346

## Abstract

This thesis aims to develop a psychologically plausible account of concepts by integrating key insights from philosophy (on the metaphysical basis for concept possession) and psychology (on the mechanisms underlying concept acquisition).

I adopt an approach known as informational atomism, developed by Jerry Fodor. Informational atomism is the conjunction of two theses: (i) *informational semantics*, according to which conceptual content is constituted exhaustively by nomological mind–world relations; and (ii) *conceptual atomism*, according to which (lexical) concepts have no internal structure.

I argue that informational semantics needs to be supplemented by allowing content-constitutive rules of inference (“meaning postulates”). This is because the content of one important class of concepts, the logical terms, is not plausibly informational. And since, it is argued, no principled distinction can be drawn between logical concepts and the rest, the problem that this raises is a general one. An immediate difficulty is that Quine’s classic arguments against the analytic/synthetic distinction suggest that there can be no principled basis for distinguishing content-constitutive rules from the rest. I show that this concern can be overcome by taking a psychological approach: there is a fact of the matter as to whether or not a particular inference is governed by a mentally-represented inference rule, albeit one that analytic philosophy does not have the resources to determine.

I then consider the implications of this approach for concept acquisition. One mechanism underlying concept acquisition is the development of perceptual detectors for the objects that we encounter. I investigate how this might work, by drawing on recent ideas in ethology on ‘learning instincts’, and recent insights into the neurological basis for perceptual learning. What emerges is a view of concept acquisition as involving a complex interplay between innate constraints and environmental input. This supports Fodor’s recent move away from radical concept nativism: concept acquisition requires innate mechanisms, but does not require that concepts themselves be innate.

# Contents

<b>ACKNOWLEDGEMENTS.....</b>	<b>5</b>
<b>1. FODOR ON CONCEPTS.....</b>	<b>6</b>
1.1. FODOR AND RADICAL CONCEPT NATIVISM .....	6
1.2. FODOR'S RETREAT FROM RADICAL NATIVISM .....	9
1.2.1. <i>Concepts and the Representational Theory of Mind</i> .....	9
1.2.2. <i>Definition theory in philosophy</i> .....	11
1.2.3. <i>Definition theory in linguistics</i> .....	15
1.2.4. <i>Prototype theory</i> .....	18
1.2.5. <i>Meaning postulates</i> .....	20
1.2.6. <i>Fodor's proposals</i> .....	21
1.3. PROBLEMS WITH FODOR'S ACCOUNT .....	25
1.3.1. <i>Analyticity and meaning postulates</i> .....	25
1.3.2. <i>Concepts for impossible entities</i> .....	27
1.3.3. <i>Acquisition</i> .....	29
<b>2. ANALYTICITY AND MEANING POSTULATES .....</b>	<b>33</b>
2.1. THE FORMAL STATUS OF MEANING POSTULATES.....	33
2.2. THE PSYCHOLOGICAL STATUS OF MEANING POSTULATES .....	38
2.3. THE LOGICAL AND NON-LOGICAL VOCABULARIES .....	44
2.4. PROBLEMS WITH INFERENTIAL ROLE SEMANTICS.....	50
2.4.1. <i>Content-constitutive inferences</i> .....	50
2.4.2. <i>Meaning holism</i> .....	51
2.4.3. <i>Molecularism</i> .....	54
2.4.4. <i>The circularity argument</i> .....	56
2.5. 'PURE' INFORMATIONAL ATOMISM .....	58
2.5.1. <i>Fodor's new proposal</i> .....	58
2.5.2. <i>Problems with 'pure' informational atomism</i> .....	60
2.6. AN ALTERNATIVE WAY FORWARD?.....	70
2.6.1. <i>The problem</i> .....	70
2.6.2. <i>Apriority and analyticity</i> .....	71
2.7. PSYCHOSEMANTIC ANALYTICITY .....	73
2.8. CONCLUSION .....	77

<b>3.</b>	<b>CONTENT .....</b>	<b>79</b>
3.1.	INTRODUCTION: WHAT IS CONCEPTUAL CONTENT? .....	79
3.2.	INFORMATIONAL SEMANTICS .....	85
3.3.	PROBLEMATIC CASES FOR INFORMATIONAL SEMANTICS .....	90
3.4.	LOGICAL CONNECTIVES AND THEIR CANONICAL INFERENCES .....	96
3.5.	ELIMINATION RULES AND MENTAL DEDUCTION .....	99
3.6.	ARE MEANING POSTULATES CONTENT CONSTITUTIVE? .....	115
	APPENDIX: THE RANGE OF AVAILABLE MEANING POSTULATES .....	124
<b>4.</b>	<b>INTUITIVE AND REFLECTIVE CONCEPTS.....</b>	<b>127</b>
4.1.	THE PROBLEM OF INCOMPLETE CONCEPTS .....	127
4.1.1.	<i>Overview of the problem</i> .....	127
4.1.2.	<i>Ontological issues</i> .....	130
4.1.3.	<i>Psychological issues</i> .....	134
4.2.	MENTALESE ORTHOGRAPHY .....	140
4.3.	WAYS OF BELIEVING .....	142
4.3.1.	<i>Intuitive and reflective beliefs</i> .....	142
4.3.2.	<i>Intuitive and reflective concepts</i> .....	153
4.4.	DIFFERENT KINDS OF DETECTOR .....	161
4.5.	CONCLUSION .....	167
<b>5.</b>	<b>ACQUISITION .....</b>	<b>171</b>
5.1.	LEARNING INSTINCTS.....	171
5.2.	INSTINCTS TO ACQUIRE CONCEPTS .....	176
5.3.	QUOTIDIAN CATEGORIZATION AND EXPERT CATEGORIZATION .....	180
5.3.1.	<i>Expert abilities</i> .....	180
5.3.2.	<i>Chicken sexing and plane spotting</i> .....	181
5.3.3.	<i>Features or gestalt?</i> .....	189
5.4.	EXPLAINING CONCEPT ACQUISITION .....	196
5.4.1.	<i>Instinct and learning in categorization</i> .....	196
5.4.2.	<i>Perceptual expertise and perceptual learning</i> .....	204
5.4.3.	<i>The acquisition of perceptual detectors</i> .....	215
5.5.	CONCLUSIONS: WHAT ARE CONCEPTS AND HOW ARE THEY ACQUIRED? .....	221
	<b>BIBLIOGRAPHY .....</b>	<b>229</b>

## Acknowledgements

My greatest intellectual debt is to my supervisor, Deirdre Wilson. Her vision and insight, and her support and advice, have inspired me in writing the thesis and have given me the momentum needed to finish it under difficult circumstances. Neil Smith provided my introduction to UCL when I came enquiring; it was he who suggested that I should talk to Deirdre if I had “interesting views on Fodor” (I doubt they were all that interesting at the time, but thanks, Deirdre, for thinking there was further potential). I would also like to thank Robyn Carston for enlightening discussions over the years, and for always being supportive of my work. Thanks to Tim Wharton for friendship, for stimulating discussions, and for epicurean delights. Thanks also to Annabel Cormack for useful comments on some draft sections. Lastly, the impact that Jerry Fodor’s writings have had on my thinking and on the development of this thesis are obvious; further elaboration should wait until after he’s examined it.

I would like to express my sincere thanks to the Association of Commonwealth Universities (and Teresa Anderson in particular). Their Commonwealth Scholarship kept the wolf from my door in the three years while I was researching the thesis in London, and they showed great flexibility and understanding at various stages when I decided to do unorthodox things that would be *prima facie* detrimental to my PhD (such as deferring my award to take up a full-time position in an unrelated field).

The writing up of this thesis was carried out while working for the International Labour Organization in Burma. I would like to thank the ILO, and Francis Maupain and Kari Tapiola personally, for their support and for convincing me that I was better off not giving up my day job (or even half of it) to finish the thesis. Friends and colleagues in Burma provided enormous support and encouragement. In particular, thanks are due to Martin and Anna, for mercy packages of food to fuel my late-night writing; Sivanka, for tennis and other welcome distractions; Beth and Reid, for pizza, pool and many discussions of academia; Ietje, for support, enthusiasm and great parties; and Charles and Elke, for Ali G and evenings by the lake.

Thanks to my parents, Max and Sally, who taught me so many things, who sacrificed a lot to put my education first, and who have seen me a lot less over the last few years than they otherwise might have. Finally, and most importantly, to Teresa for her love and friendship and endless support. Evenings, weekends and holidays have seen me hunched over a computer muttering about meaning postulates; thank you for understanding, and for sacrificing a lot to see this finished. Annika, having fortunately not yet developed episodic memory, will be oblivious to my early lapses of fatherhood; but she provided the final impetus to finish the PhD.



# 1. Fodor on concepts

## 1.1. Fodor and radical concept nativism

Concepts may be thought of as coming in two kinds: *lexical* and *phrasal*. Lexical concepts are (approximately) those corresponding to monomorphemic lexical items.<sup>1</sup> Phrasal concepts are all the rest (that is, those corresponding to morphologically complex lexical items and phrases). Classical theories of concepts also come in two kinds: *empiricist* and *nativist*. Empiricists and nativists agree about the nature of phrasal concepts, but they disagree on the nature of lexical concepts.

For both empiricists and nativists, phrasal concepts are typically complex (they have internal structure just as the phrases corresponding to them do).<sup>2</sup> Nativists, following the analogy of the lexical/phrasal distinction, consider lexical concepts to be typically atomic (they have no internal structure, just as the corresponding lexical items do not).<sup>3</sup> Empiricists, on the other hand, consider that in addition to phrasal concepts, the majority of lexical concepts are also complex. For empiricists, lexical concepts are typically definable and have a complex structure corresponding to their phrasal definition (so the concept corresponding to the word ‘bachelor’ would be UNMARRIED MAN). The small set of lexical concepts that cannot be defined have no internal structure and are said to be *primitive*. The empiricists generally consider the primitive concepts to be exclusively sensory (apart from a small number of logical connectives); nativists clearly have to allow for other kinds of primitive concept, since for them nearly all lexical concepts are primitive.

---

<sup>1</sup> We say ‘approximately’ since there is of course some cross-linguistic variation in the set of monomorphemic lexical items. The distinction between lexical and phrasal concepts is introduced for heuristic purposes only, so there is no need to be too concerned with a precise characterization of the term ‘lexical concept’.

<sup>2</sup> Both allow that some phrasal concepts are not complex, as in the case of idioms (at least on some accounts).

<sup>3</sup> Nativists can allow that *some* lexical concepts are complex, but if they do, most nativists would claim that it is not very many (although, as Fodor 1981: 281 points out, it is in principle open to the nativist to claim that the primitive/complex distinction is far from coextensive with the lexical/phrasal distinction).

Next consider concept acquisition. It has been common ground between classical empiricists and nativists that learning a concept is an inductive process. Essentially, learning the concept XYZ involves formulating and testing hypotheses of the form ‘ $x$  is XYZ iff  $x$  is ...’, where ‘...’ is a specification of what it is to be an XYZ. While this is reasonable for complex or definable concepts (for example, ‘ $x$  is a BACHELOR iff  $x$  is an UNMARRIED MAN’), it clearly cannot account for the acquisition of primitive or undefinable concepts (since we would already have to have the concept we were trying to acquire in order to frame the hypothesis: consider ‘ $x$  is RED iff  $x$  is RED’). It follows, for both empiricists and nativists, that primitive concepts must be innate—or at least unlearned in the sense described above. The difference, of course, is in the nature and number of primitive concepts. The empiricist claim is that a small number of sensory concepts (plus a few logical connectives) are innate. Nativists, on the other hand, seem to be forced to adopt a rather radical position: that virtually all lexical concepts are innate. This position is known as ‘radical concept nativism’. Nativists do not claim that experience plays no role in concept acquisition; but they see concepts as being *triggered* by experience in a brute-causal way (rather than learned from experience in a rational-inductive way).

In an influential 1981 paper, “The present status of the innateness controversy”, Jerry Fodor presented arguments against the classical empiricist position on concepts. He pointed out that the available evidence strongly suggested that the majority of lexical concepts have no internal structure. This was the reason, in his view, for the failure of analytic philosophy. It also explained why it had proved almost impossible to come up with plausible examples of definitions (let alone, as Empiricism would require, definitions of non-sensory terms in a purely sensory vocabulary).

As a concrete example, Fodor considered a proposed definition of the transitive verb ‘paint’: ‘ $x$  covers the surface of  $y$  with paint<sub>n</sub>’. Prima facie, this seemed to be a

plausible definition (and had been presented as such in the literature<sup>4</sup>). But Fodor pointed out that on closer consideration the definition was inadequate. Suppose that a paint factory explodes and covers bystanders with paint. This seems to fit the proposed definition, but the paint factory clearly cannot be said to have painted the bystanders. This suggests the extra condition that *x* must be an agent should be added to the putative definition for 'paint<sub>n</sub>'. But Fodor pointed out that this new definition was also inadequate. When Michelangelo covered the ceiling of the Sistine Chapel with paint, it is doubtful whether he could be said to have been painting the ceiling (rather, he was painting a picture *on* the ceiling). One could supplement the definition again, perhaps by stipulating that covering the surface with paint be the primary intention of the act (rather than, say, a consequence of the primary intention of putting a picture on the surface), but this leaves PRIMARY INTENTION OF AN ACT as a completely unanalysed part of the proposed definition, to which learners of 'paint<sub>n</sub>' would have to have prior access. This somewhat implausible definition is still not adequate, however. For when Michelangelo dipped his brush into a pot of paint, he thereby covered the surface of the brush with paint and moreover his primary intention was that it should be so covered. But, as Fodor pointed out, we would not say that he painted his paintbrush.

Fodor also noted that the psycholinguistic evidence was suggestive of most lexical concepts being non-complex. Experiments that had been carried out indicated that understanding putatively complex lexical concepts (such as BACHELOR) was no slower than understanding primitive lexical items (such as MAN). This evidence, and related arguments such as the fact that putatively complex concepts like BACHELOR that included a negative element (UNMARRIED) did not give rise to the processing difficulties that negative items standardly do, were discussed in detail in Fodor et al. (1980).

---

<sup>4</sup> Miller (1978). Similar examples abounded in the generative semantics literature (see, for example, the treatment of causal verbs such as 'break' in Lakoff & Ross 1976, or the treatment of 'kill' in McCawley 1968; for an overview of this literature, see J. D. Fodor 1977).

There remained a puzzle for Fodor. If concepts were triggered by experience, rather than constructed out of other more basic concepts, then it would appear that the order in which we acquire concepts is essentially arbitrary.<sup>5</sup> Fodor's proposal was that there may be a hierarchy of triggering ("the triggering structure of the mind is layered") which predicted the observed order of concept acquisition. On this view, basic-level concepts are acquired early because they are in the first "layer", whereas other levels of concepts are in more remote layers and therefore acquired relatively late. This proposal was little more than a restatement of the problem, however.

## **1.2. Fodor's retreat from radical nativism**

### ***1.2.1. Concepts and the Representational Theory of Mind***

In his book *Concepts: Where Cognitive Science Went Wrong* (1998a), Fodor presents further arguments for conceptual atomism, but he retreats from the radically nativist position that he set out in "The present status of the innateness controversy" (1981). The position that Fodor (1998a) sets out will form much of the background to this thesis, and so I intend to summarise it in some detail in what follows.

Fodor (1998a) starts with a defence of the Representational Theory of Mind, which he takes to consist in the conjunction of the following five theses:

*First thesis.* The laws that psychological explanations invoke typically express causal relations among mental states (beliefs, desires, and so on) that are picked out by reference to their contents.

*Second thesis.* Mental representations are the primitive bearers of intentional content (upon which the intentionality of propositional attitudes and natural language is parasitic).

*Third thesis.* Thinking is computation. Mental processes are (content-respecting) causal relations among mental representations.

---

<sup>5</sup> Though, of course, factors such as environmental frequency will play a role.

*Fourth thesis.* Meaning is information (what Fodor calls “informational semantics”). Mental representations get their content from their causal-nomological relations with the things that fall under them. (So it’s something about its causal-nomological relations with tigers that gives the concept TIGER its content.) This means that coreferential concept representations must be synonymous. But although the content of coreferential concepts must be the same, Fodor is prepared to allow that they are different concepts. That is, there is more to *concept* individuation than *content* individuation.

*Fifth thesis.* Whatever distinguishes coextensive concepts is ‘in the head’. Fodor follows Frege in proposing that what distinguishes coextensive concepts is the ‘mode of presentation’. However, given that informational semantics implies that coextensive concepts are synonymous (fourth thesis), he must reject the standard Fregean position that modes of presentation are *senses* (or meanings). He assumes that the mode of presentation of a concept is the mental representation itself (that is, an expression of Mentalese).

Having thus set the scene, Fodor goes on to set out five “non-negotiable conditions on a theory of concepts” in order that the theory be compatible with the Representational Theory of Mind:

*Condition 1.* Concepts are mental particulars, which function as mental causes and effects. (This condition is entailed by the Representational Theory of Mind.)

*Condition 2.* Concepts are categories, and things in the world fall under them.

*Condition 3.* Concepts are compositional. They are the constituents of thoughts and of other (complex) concepts.

*Condition 4.* Many concepts (that is, those that are not primitive) must be learned.

*Condition 5.* Concepts are public (that is, people can and do share them). This is required if intensional generalisations are to have any explanatory power, for if each person's WATER concept is slightly different then the generalisation "thirsty people seek water" cannot apply to all of them. Fodor argues that it is not possible to avoid this by resorting to some notion of concept *similarity*, since accounts of concept similarity invariably presuppose an account of concept identity.

In the remainder of the book, Fodor discusses why current theories fail to meet the five conditions that he has set out. He considers two current theories in detail (definition theory and prototype theory), and argues that they both fail these constraints because of a common assumption: that the contents of primitive concepts are constituted, at least in part, by their inferential relations.

### ***1.2.2. Definition theory in philosophy***

A once-popular theory held that concepts were definitions, and that having a concept was being disposed to draw the inferences that define it. On this view, just as the word "bachelor" can supposedly be defined by the phrase "unmarried man", so the concept BACHELOR supposedly has the constituent structure UNMARRIED MAN. This theory makes two main assumptions: that many words are definable, and that in many cases the mental representations corresponding to these definable words have the same structure as the definition.<sup>6</sup>

---

<sup>6</sup> As Fodor points out (1998a: 41, fn. 1; cf. 2003a: 64), when we use phrases such as "the concept BACHELOR" it is important to be clear whether BACHELOR is to be read as the name of the concept or its structural description. Here I am following Fodor's practice that words appearing in small capitals are names of concepts which leave open the question of what the structure of the mental representation is. So, for example, on the definitional view the mental representation corresponding to BACHELOR would have been UNMARRIED MAN—or in fact something far more complex, since (particularly for empiricists) UNMARRIED and MAN would themselves have been definable. Empiricists would have needed to make some provision for mental abbreviation, to avoid having to posit mental

Fodor made forceful arguments against this theory in earlier work (for example, Fodor 1981), and few cognitive scientists now consider that definitions have an important role to play in theories of meaning. Fodor provides three reasons for the collapse of definition theory:

—Despite quite a lot of effort on the part of definition theorists, almost all attempts to give definitions for words have failed, strongly suggesting that most words can't be defined.

—Definitions don't seem to have any psychological reality: understanding sentences with putatively complex words has been demonstrated to be no more difficult than understanding sentences with non-complex words<sup>7</sup>; neither do definitions seem to play any role in concept acquisition or reasoning.

—As Quine's (1953a) arguments against analyticity showed, there doesn't seem to be a principled way to distinguish defining inferences from the rest. This is an instance of a more general problem for any theory that sees the content of a concept as being constituted by its inferential role.

As already mentioned, Fodor considers that current theories of concepts are all unsatisfactory because they fail on this last point, so it is worth looking at his position here in more detail. Quine's main argument against the possibility of drawing a principled analytic/synthetic distinction was precisely that it has proved impossible for anyone to draw a principled distinction between those inferences that are content constitutive and those that are not. Moreover, Fodor thinks that

---

representations consisting of extremely long and computationally difficult conjuncts of sensory primitives.

<sup>7</sup> This is supposed to be the case even for lexical concepts that *do* have reasonably plausible definitions, such as BACHELOR, TRIANGLE, FATHER and TUESDAY. So although being a bachelor entails being unmarried (and being a man), Fodor takes this to be a fact about the property *bachelorhood* rather than a fact about the concept BACHELOR.

informational semantics can explain why Quine was right about this: informational semantics is inherently atomistic, and so denies that content is constituted by *any* inferential relations.

One problem that Quine had battled with, however, was why, if there was no analytic/synthetic distinction, we have such strong intuitions that certain propositions (for example, ‘no bachelors are married’ or ‘Tuesdays precede Wednesdays’) are analytic.<sup>8</sup> Fodor thinks that he can explain these intuitions as well. The standard (Quinean) explanation relies on appeals to centrality. Some of our beliefs are more central than others, in the sense that revising them would cause a greater disturbance to our belief system as a whole. For example, fundamental logico-mathematical beliefs are highly central in this sense—revising them would force us to revise a large number of other beliefs, and so they tend to be immune from revision. Basic physical laws are also central, although less so than logico-mathematical statements. Quine proposed that the strong intuitions we have that some propositions are analytic might actually be intuitions of centrality. This would explain our strong intuitions about the analyticity of logical statements such as ‘no unmarried men are married’. And if we have the intuition that the proposition ‘ $F = Ma$ ’ expressing the relation between force, mass and acceleration is analytic then on Quine’s account perhaps this is because it is central to our scientific views about the behaviour of middle-sized objects. This might be a reasonable account for many propositions, but as Putnam (1983) pointed out, it doesn’t always work quite so well. There seem to be propositions that intuition tells us are analytic, but which do not play any central role in our belief systems. For example, ‘Tuesdays come before Wednesdays’ would seem to be a clear case of an intuitively analytic proposition, but it would be difficult to argue that such propositions play any central role in our belief systems.

---

<sup>8</sup> Strictly speaking, of course, the intuitions in question are not that such propositions are analytic, but that they are true ‘in some special way’, perhaps. What exactly they are intuitions *of* is precisely the point at issue.



Fodor's strategy for dealing with intuitions of analyticity is as follows. He continues to adopt Quine's explanation for those cases where centrality is a plausible explanation (that is, for cases such as ' $F = Ma$ ' which are plausibly seen as being central to our belief systems). For those cases where it is not (cases such as 'Tuesdays come before Wednesdays' or 'no bachelors are married'), he suggests that informational semantics can provide an explanation. First, he notes that informational semantics only rules out content-constitutive conceptual connections: it denies, for example, that it's necessary to have ANIMAL in order to have CAT. It doesn't deny that there's a necessary connection between *cathood* and *animalhood*, and therefore doesn't rule out intuitions of such a necessary connection. Fodor then notes that if cats reliably cause tokenings of CAT, as informational semantics requires, there must be some (causal) mechanism that sustains this relation.<sup>9</sup> Informational semantics is not concerned with what this relation is, merely that there is one (so visual perception may play an important role in mediating between *cat* and CAT in many cases, but this cannot be the case for blind people, or in a range of situations where there is a more indirect link, such as forensic detection of cats). To explain intuitions of analyticity in non-central cases, Fodor co-opts the idea of 'one-criterion concepts' put forward by Putnam (1983). For some concepts, there are lots of (independent) ways to determine if the concept applies, whereas for other concepts there is just one (independent) way. So, there are plenty of ways to determine whether WATER applies to a substance, such as performing a chemical analysis, tasting it, and so on, whereas there is only one way to tell whether BACHELOR applies to a person, viz. to determine whether they are an unmarried man. Fodor proposes that if one thinks that the only way to tell whether a concept *C* applies is to determine whether concept *C\** applies, then this will give rise to the intuition that *C* and *C\** are conceptually connected and that the proposition

---

<sup>9</sup> Fodor refers to this mechanism as "semantic access". I will generally follow Margolis (1998) and Laurence & Margolis (1999a) in using the term "sustaining mechanism".

'Cs are C\*s' is analytic. So intuitions of analyticity are actually intuitions of something else: either centrality, or one-criterionhood.

The question immediately arises, of course, whether *all* our intuitions of analyticity can be explained by these two accounts. Fodor considers that his one-criterionhood account works well in just those cases where appealing to centrality doesn't. However, there would seem to be intuitively analytic propositions that are difficult to explain on either account, such as: 'cats are animals' or 'red is a colour'. Fodor (1998a: 84) discusses such propositions and suggests, in fact, that there is no strong intuition that they are analytic; as noted above, according to Fodor there may be a necessary connection between *cathood* and *animalhood*, about which we may have intuitions, but he claims that the intuition here is not one of analyticity. One of the problems is that it's difficult to regain our innocence, as it were, and recover our 'pure' intuitions of analyticity; we often find we have to consult our theoretical commitments to check what our intuition says. It has been remarked by Gilbert Harman that "someone could be taught to make the analytic-synthetic distinction only by being taught a rather substantial theory, a theory including such principles as that meaning can make something true and that knowledge of meaning can give knowledge of truth".<sup>10</sup> We will discuss the question of analyticity and associated intuitions in more detail in chapter 2 (see also Laurence & Margolis 2003a for relevant discussion).

### ***1.2.3. Definition theory in linguistics***

Although there are persuasive arguments against the theory of definitions, reviewed above, Fodor notes there is some evidence from the field of lexical semantics that there exists a level of linguistic analysis at which many words are represented by decompositions. Lexical semanticists are typically of the view that these decompositions fall short of being true definitions, in the sense that they capture only 'core meaning' rather than representing all aspects of meaning. However, Fodor

---

<sup>10</sup> "Doubts about conceptual analysis", Ms., p. 5, quoted in Boghossian (1997: 337).

argues that, particularly given that no explication of ‘core meaning’ is provided, these decompositions face similar problems to definitions. He looks in particular at influential arguments by two authors, Jackendoff and Pinker. Here I will summarise his discussion of Jackendoff.

Jackendoff (1992) discusses the example of the verb ‘keep’. He notes that ‘keep’ is polysemous in (1)–(4) below, rather than ambiguous, since intuition tells us that that it is the same word that occurs in each of these sentences:

- (1) Harry kept the bird in the cage [semantic field: spatial location and motion]
- (2) Susan kept the money [semantic field: possession]
- (3) Sam kept the crowd happy [semantic field: ascription of properties]
- (4) Let’s keep the trip on Saturday [semantic field: scheduling of activities]

But there is also the intuition that the sense of ‘keep’ is different in each sentence: the relation expressed between Susan and the money in (2) isn’t exactly the same as the relation expressed between Sam and the crowd in (3). Jackendoff’s claim is that we can reconcile these intuitions if we say that ‘keep’ expresses ‘the causation of a state that endures over a period of time’. This accounts for our intuition that ‘keep’ is univocal. The intuition that there are different senses of ‘keep’ is explained by differences among the semantic fields, each of which ‘has its own particular inferential patterns’.

Fodor’s response is that this account of polysemy takes for granted a theoretical vocabulary whose own semantics is unspecified in crucial respects. Jackendoff’s proposal is that whatever semantic field it occurs in, ‘keep’ always expresses the concept CAUSE A STATE THAT ENDURES OVER TIME. But this would explain the intuitive univocality of ‘keep’ only on the assumption that the constituent concepts CAUSE, STATE, ENDURE and TIME and so on are themselves univocal across semantic fields. Fodor therefore asks us to consider whether CAUSE is univocal in CAUSE THE MONEY TO BE IN SUSAN’S POCKET and CAUSE THE CROWD TO BE HAPPY. He suggests that Jackendoff is in trouble whichever answer is given.

For if it is accepted that CAUSE is polysemic, then ‘ $\varphi$  CAUSE  $\psi$ ’ is itself polysemic, so the assumption that ‘keep’ always means ‘ $\varphi$  CAUSE  $\psi$ ’ doesn’t explain why ‘keep’ is intuitively univocal. If, on the other hand, CAUSE is univocal, how is this fact to be explained? There would seem to be two possibilities: either CAUSE has the definition ‘ $\sigma X \tau$ ’ where ‘ $X$ ’ is univocal across semantic fields, or else CAUSE is primitive and always means *cause*. The first of these possibilities is unappealing since it leads to an infinite regress. The second possibility is an admission that it is not necessary to have a decomposition/definition for a word (concept) in order to explain the fact that it is univocal across semantic fields, since CAUSE would be an example to the contrary. In which case why not just give the same account for ‘keep’, and say that it is univocal because it always means *keep*? For Fodor, the fact that the relation expressed by ‘keep’ in different situations can be explained by differences in the relata (money is very different from the crowd’s being happy); it is not necessary to also posit a difference in meaning for ‘keep’. Fodor therefore explains polysemy by eliminating it: for him there is no such thing.

Fodor discusses one potential difficulty with his position. Consider cases where language *A* has a single unambiguous word ( $\alpha$ ), the translation of which in language *B* is one of two words ( $\beta, \beta'$ ), depending on the context. In such cases, isn’t  $\alpha$  ipso facto polysemous? Fodor thinks not. He takes the example of the English words ‘spoiled’ and ‘addled’, both of which are taken to mean *spoiled*, but one of which is used only of eggs (discussed in Quine 1973/1976). Fodor then asks us to consider a hypothetical second language having only a single word, ‘spoilissimoed’, which means *spoiled* and is used generally (including for eggs). He suggests that the right way to look at this situation is not to invoke polysemy, but to say that all these words mean *spoiled* (and are therefore synonymous). The difference between the languages is that one, but not the other, has a word meaning *spoiled* but which is context-restricted to eggs (that is, the possession condition for which includes having the concept EGG). We will return to consideration of this argument in §2.5.2.

#### 1.2.4. *Prototype theory*

Whereas the definition theory of concepts claims that complex concepts typically entail their constituents (so *is a bachelor* entails *is a man*, for example), prototype theories of concepts claim that it is the reliability of inferences that determines their relevance to concept individuation (so being able to fly isn't a necessary condition for being a bird, but it is a reliable property of birds). On this view, a concept is a bundle of (not necessary but) statistically reliable features. This is a version of the thesis that concepts are individuated by their inferential relations; in the present case, these inferences are regarded as statistical.

While definition theory suffered from the problem that there didn't seem to be any definitions, the evidence for prototype effects is well-documented and strong. Fodor argues, however, that there is a fundamental flaw with prototype theory: concepts are compositional, but prototypes are notoriously non-compositional, so concepts just cannot be prototypes. Fodor gives a number of arguments why concepts must be compositional:<sup>11</sup>

—*The productivity argument.* There are an infinite number of concepts that we can entertain, but our brains have only finite capacity. This situation can be explained if (and possibly only if) complex concepts are individuated by their contents, which are inherited from the contents of their primitive constituents plus their mode of combination (that is, their syntax).<sup>12</sup>

—*The systematicity argument.* Compositionality explains why cognition tends to be systematic. If one has the capacity to think ' $P \supset Q$ ', then one also has the capacity to think ' $Q \supset P$ ', but one does not necessarily have the capacity to think ' $R \supset Q$ '. Compositionality explains this because on

---

<sup>11</sup> Fodor notes (1998a: 94) that a rigorous characterization of compositionality is lacking. Intuitively, however, the claim is that the syntax and the content of a complex concept is determined by the syntax and the content of its constituents, other than for a finite number of idiomatic expressions.

<sup>12</sup> As Fodor points out, connectionists don't necessarily accept that we can, even in principle, entertain an infinite number of concepts. But he considers that the systematicity argument is still fatal.

a compositional account the propositions ' $P \supset Q$ ' and ' $Q \supset P$ ' share the same constituents, whereas the proposition ' $R \supset Q$ ' does not.

—*Other arguments.* The best argument for compositionality, in Fodor's view, is that evidence for it is ubiquitous. He gives the example of definite descriptions, which make it possible to pick out entities by reference to their properties. But this requires compositionality: the definite description 'the brown cow' (and the concept it expresses) only picks out a certain cow with the property *brown* because of compositionality. He compares this to the situation with names, which aren't compositional, and where 'the Iron Duke' does not pick out a certain duke with the property *iron*.

Fodor's argument is that prototypes fail compositionality, and hence that concepts cannot be prototypes. He presents three problems that prototypes face in passing the compositionality test:

—*The UNCAT problem.* There are indefinitely many complex concepts that do not have any prototype (and therefore fail compositionality). Take, for example, the concept NOT A CAT. This has no prototype (since there is nothing that non-cats tend to have in common, apart from not being cats).

—*The PET FISH problem.* There are indefinitely many complex concepts whose prototypes are not composed from the prototypes of their constituents. According to prototype theory, an entity falls under a concept to the extent that it is similar to the concept's exemplar. Thus, an entity's similarity to the exemplar for a complex concept has to be determined by its similarity to the exemplars for its constituent concepts. But this is not generally the case. For example, a goldfish is not a prototypical example of a fish, nor a prototypical example of a pet, but is a prototypical example of a pet fish.

—*The problem of inferential roles.* Quite apart from these two specific problems, Fodor suggests that there is a convincing general argument why concepts can't be prototypes. The reason is that prototype theories of conceptual content are a version of 'inferential role semantics' (since they hold that a particular set of inferences—viz., the statistically reliable ones—are content constitutive). The problem is that inferential roles aren't compositional. Take the complex concept BROWN COW. There are lots of inferences (including statistically reliable ones) that could hold of brown cows that do not hold of brown things in general or of cows in general. This shows that the *PET FISH* problem is not isolated, but is a special case of a more general problem for inferential role theories.<sup>13</sup>

#### 1.2.5. *Meaning postulates*

Both the definition and prototype theories of concepts treat lexical concepts as structurally complex. But the idea that inferences are content constitutive doesn't entail that lexical concepts are complex. Even if one adopts an atomistic view of lexical concepts, one can still claim that their content is constituted by inferences.

This has been proposed by some philosophers<sup>14</sup> as a solution to the 'residuum problem'. This problem arises for certain kinds of entailment in the definition theory of concepts. Consider the entailment *bachelor*  $\Rightarrow$  *unmarried*. This entailment falls naturally out of the definition theory by an analogue of simplification of conjunction: *bachelor* entails *unmarried* because the definition of 'bachelor' is *man & unmarried*. But consider an entailment such as *red*  $\Rightarrow$  *colour*. Along the lines of the previous example, this should be a simplification of the definition of 'red' as *colour & X*. But it is not at all clear what the 'X' could stand for: what concept (other than RED itself)

---

<sup>13</sup> The only exception is the version of inferential role semantics according to which it's the definitional inferences that are content constitutive. Definitional inferences are (by definition) compositional. The problem, of course, is that there don't appear to be any definitions (see §1.2.2 above).

<sup>14</sup> Most notably, by Fodor himself in Fodor (1975) and J. D. Fodor et al. (1975: 519), although the term itself comes from Carnap (1952, 1956), who used it in a somewhat different sense.

could combine with COLOUR to give RED? The same is true for *dog*  $\Rightarrow$  *animal*, and indefinitely many other cases.

This problem can be accommodated if one assumes that not all content-constitutive inferences need arise from definitions: some may be basic (or stipulated). Such inferences are sometimes called ‘meaning postulates’. Introducing meaning postulates allows atomism to be adopted, whilst maintaining the notion of analyticity. But Fodor points out that such an account can no longer explain *why* certain inferences are analytic. He suggests that no principled answer can be given to the question “When is an entailment merely necessary, and when is it a meaning postulate (that is, content constitutive)?”. We consider the status of meaning postulates in more detail in chapter 2.

#### ***1.2.6. Fodor’s proposals***

Having discussed in detail what concepts *aren’t*, Fodor then goes on to present some positive proposals. He puts forward the doctrine of ‘informational atomism’, which has two parts:

- Informational semantics*. Content is constituted by some kind of nomic, mind–world relation (and therefore having a concept is constituted, at least in part, by being in some such relation).
- Conceptual atomism*. Most lexical concepts have no internal structure (they are atomic).

Fodor considers two potential objections to informational atomism in detail. The first objection is that informational atomism implies radical concept nativism, which most people have difficulty in accepting. The second objection is that if content is constituted by nomic mind–world relations, then there would have to be laws about everything for which we have (atomic) concepts; this seems implausible.



### *Concept nativism*

There is a plausible ‘standard argument’ that informational atomism implies radical concept nativism.<sup>15</sup> This standard argument assumes that the only method for concept learning is induction: devising and testing hypotheses about what the property is that entities must have in order for them to fall under a particular concept. Clearly, atomic concepts can’t themselves be learned in this way, since unless one already has the concept, there is no way of mentally representing the hypothesis. The conclusion is that atomic concepts must be innate, and that if (as informational atomism posits) most lexical concepts are atomic, then most lexical concepts must be innate.

One possible reply is that the inductive account of concept acquisition is plausible only if one assumes that concept possession is a kind of ‘knowledge that’. This is plausible on a definition account of concepts (having a concept is knowing a certain definition), but if concepts are mind–world relations, it’s plausible that concept possession is more like ‘knowledge how’. Concept acquisition could then arguably be learning how, which isn’t necessarily inductive. This line of thought is appealing, but ultimately Fodor finds it unconvincing. It’s not after all clear that learning how doesn’t depend on learning that (for example, learning *how* to speak a language requires learning *that* the language has the particular grammar that it does). Concept possession also involves certain perceptual and inferential skills which similarly seem to require *knowledge that*. For example, it seems that one can’t identify a dog by its barking unless one *knows that* dogs bark.

But although he finds the ‘knowledge *how*’ argument unconvincing, Fodor believes that informational atomism can offer an alternative to concept acquisition by induction and thereby avoid radical concept nativism. Since he previously appeared to embrace radical concept nativism, Fodor first explains why it is that he now wishes to avoid it. On a radically nativist account, such as that put forward in Fodor (1981), innate concepts are ‘triggered’ by experience of the entities falling

---

<sup>15</sup> This argument was first put forward by Fodor himself, in *The Language of Thought* (1975).

under them. But this can't explain what he calls the 'doorknob/DOORKNOB problem': why it is typically experiences of doorknobs (say), rather than other kinds of experiences, that trigger DOORKNOB. An inductive account solves this problem trivially, since the inductive process is evidential, and requires experiences of instances for its evidence. But it's not at all clear how a process of triggering can solve the problem.<sup>16</sup>

Fodor proposes an alternative to inductive learning that avoids radical concept nativism and solves the doorknob/DOORKNOB problem. The alternative mechanism of concept acquisition that he offers is a (non-inductive, 'brute causal') process of getting *nomologically locked* to the property that the concept expresses. Consider the following. If 'doorknob' doesn't have a definition, then what makes a doorknob a doorknob? Doorknobs don't (like natural kinds) have a hidden essence, and neither can *doorknob* be a metaphysically primitive property like *charge* and *spin*. Fodor's answer is that *doorknob* is an 'appearance property', like *red*. For 'red' can't be defined, and *redness* doesn't have a real essence, nor is it metaphysically primitive. Rather, *redness* is an appearance property: the property of striking minds like ours as being red. Fodor's proposal is that appearance properties need not be sensory, and so *doorknob* is the property things have of striking minds like ours as being doorknobs. This solves the doorknob/DOORKNOB problem, because if being a doorknob is constituted by how things strike minds like ours, then there is a metaphysically necessary connection between DOORKNOBS and doorknobs. It's experience of doorknobs that causes us to lock to DOORKNOB because the property *doorknob* just is the property that minds like ours lock to from experiences of the doorknob stereotype.<sup>17</sup> Thus, acquiring DOORKNOB requires only that we have the kinds of

---

<sup>16</sup> Fodor briefly considers the possibility that evolution could provide a solution: we have evolved in such a way that it's the right kind of experiences that act as triggers. But he rejects this possibility, since he considers that the only plausible evolved mechanism for ensuring this would be inductive learning.

<sup>17</sup> Fodor notes, however, that this relation is contingent: in some cases we may lock to a different property on exposure to the doorknob stereotype (for example, *belongs to Jones*).

minds that lock to *doorknob* in the presence of stereotypical doorknobs; we need only to posit innate mechanisms (the sensorium, etc.), not innate concepts. Fodor thinks that this story about the mind-dependence of doorknobs is true not only of artefact concepts, but of *all* concepts other than logico-mathematical concepts (which at the time he took to be constituted by meaning postulates) and concepts of natural kinds (on which more below).

### *Concepts of natural kinds*

Fodor then considers a second possible objection to informational atomism. If content is constituted by nomic mind–world relations, this implies that there must be laws about everything that we have concepts of. But it seems implausible that there could be laws about doorknobs and the like. Fodor notes that the law that experiences of doorknobs (say) cause us to lock to *doorknob* is not really a law about doorknobs at all; it is really a law about us. This is unproblematic, because we (unlike doorknobs) are a natural kind.<sup>18</sup>

Of course, not all of our concepts express properties that are mind-dependent like *doorknob*. The properties expressed by natural kind concepts are not mind-dependent in this way: the concept WATER expresses the property *water*, which is the property of *being H<sub>2</sub>O* rather than the property of striking minds like ours in a certain way. Fodor points out that there are in principle, however, two ways in which one could have concepts of natural kinds. One could have them ‘innocently’, without any commitment to an underlying real essence (in which case they are mind-dependent). Or one could have them ‘as such’, taking them to be concepts of natural kinds having an underlying real essence that is causally responsible for their superficial properties. He considers that ‘innocent’ concepts of natural kinds are historically, ontogenetically and phylogenetically prior to concepts of natural kinds ‘as such’.

---

<sup>18</sup> Fodor points out that being mind-dependent is not incompatible with being *real*, so he can still be a metaphysical realist (about doorknobs, and everything else). In any case, many concepts (in particular, concepts of natural kinds) are not mind-dependent, and so this issue does not arise in those cases.

Innocent concepts of natural kinds can be acquired in the same way as DOORKNOB, so all that remains to account for the acquisition of concepts of natural kinds ‘as such’ is to explain how such concepts can emerge once we have acquired the innocent variety. And for Fodor, this basically requires ‘doing the science’: constructing a true theory<sup>19</sup> of what the essence of the natural kind in question is, and convincing oneself of its truth. Then, one will be disposed to consider an entity to fall under a particular concept only when the theory says that it does.

### **1.3. Problems with Fodor’s account**

#### ***1.3.1. Analyticity and meaning postulates***

Introducing meaning postulates allows an atomist to maintain the notion that content can be constituted in part by inferential relations. As we saw above (§1.2.5), however, Fodor now rejects meaning postulates. His reason for this rejection is that, given Quine’s powerful arguments against the analytic/synthetic distinction, which he is greatly influenced by, he feels that no principled answer can be given to the question “when is an inference merely necessary, and when is it content constitutive?”.

But the logic of Fodor’s position relies on the assumption that content-constitutive inferences and analytic inferences are one and the same. If we could specify a property other than analyticity which could be used to determine which inferences were content constitutive, we could thereby sidestep Quinean arguments against meaning postulates. And indeed, it has been proposed that the notions of content constitutivity and analyticity do pull apart (see, for example, Boghossian 1993, 1994, 1997). The reason for this is that it is perfectly possible for an inference to be content constitutive for a subject if the subject regards the inference as valid. But this in no way requires that the inference actually *is* valid, and therefore does not require that the inference is analytic. What we need to provide is a suitable

---

<sup>19</sup> I assume that Fodor is speaking loosely here when he speaks of constructing a ‘true theory’; presumably it would be enough for us to construct a theory to the best explanation.

characterization of ‘content-constitutive inference’. Meaning postulates conceived of in a psychological sense, as mentally-represented inference rules, can provide just such a characterization.

Once upon a time, when I started writing this thesis, I had a second argument in support of meaning postulates. The argument went something like this. Fodor could not, I noted, do without meaning postulates completely. For he needed them to account for the content of the logical vocabulary (AND, OR, NOT and so on), which almost everyone (including Fodor at that time) agreed was to be done by meaning postulates. This would not have been a problem if a clear distinction could be drawn between the logical and non-logical vocabularies. But it cannot. Many words with referential content also enter into content-constitutive logical relations: for example, ‘raw’ entails ‘NOT cooked’, ‘sibling’ entails ‘brother OR sister’, and so on. The only way to capture these logical relations is with meaning postulates. This was one more reason, I noted, to welcome any atomistic account of concepts that can incorporate meaning postulates while avoiding Quine’s concerns. However, Fodor has changed his position in the meantime (2004a, 2004b), and now eschews content-constitutive meaning postulates *even for the logical vocabulary*. This brings him one step closer to a ‘pure’ informational atomism, according to which all (lexical) concepts are unstructured, and the possession of one concept is metaphysically independent of the possession of any other concepts.

In chapter 2, I argue that there is in fact a range of cases where ‘pure’ informational atomism is inadequate, and that in many of these cases, allowing a combination of informational semantics together with content-constitutive meaning postulates can offer a solution. More broadly, chapter 2 defends the notion of meaning postulates against Quinean attack, and argues that we need meaning postulates in order to be able to characterize conceptual content, for both the logical and non-logical vocabularies. It also demonstrates that Fodor’s key concern—that there is no principled way to tell when an inference is merely necessary and when it is content constitutive—can be addressed by considering in detail the psychological

processes associated with concept acquisition and deployment. It may be tempting at this point for Fodor to accept the need for meaning postulates to support mental inference, say, but to deny that they are constitutive of content (a position adopted by de Almeida 1999; cf. Fodor 2004a, 2004b). Chapter 3 examines this possibility in detail, and concludes that meaning postulates cannot *but* be constitutive of content. I also conclude that the ‘circularity argument’ against an inferential role semantics for the logical vocabulary, recently put forward by Fodor, does not apply to the specific formulation that I present.

### 1.3.2. *Concepts for impossible entities*

In their review of Fodor (1998a), Keil & Wilson (2000: 316) consider the following case:

A biologist can tell me that he knows there must have been a mammal, which he calls a ‘schmoo’, that existed in a certain niche 20 million years ago because of some unique mammalian genetic fragment found in amber. The fragment is just enough to indicate that it was both a mammal and different from all other known mammals, but no one has the faintest idea what sort of mammal it was. We all have the concept of the ‘schmoo’ but could never identify one.

They make the point that although we (allegedly) have the concept SCHMOO, we cannot have acquired it by locking onto the property *schmoo*, since we don’t know what that property is (and nor does anyone else, including schmoo-experts, to the extent that there could be such people). In fact, however, the informational atomist has a straightforward response here, as Fodor (2000a) points out. The response has two parts.

*First part (being the answer to a metaphysical question):* How can we be locked to *schmoo* when there aren’t any schmoos? It is important to underline that being locked to a property requires that the property exists, but does not require that it is instantiated. If one is an ontological realist (as Fodor is), then it is fine for there to be

such a property as *dodo*, say, even though it is not instantiated. It's therefore perfectly okay for us to have a concept DODO. Likewise, it's perfectly okay for us to have the concept SCHMOO, even if the property *schmoo* is uninstantiated.

*Second part (being the answer to a psychological question):* Yes, but how could we *get ourselves* locked to an uninstantiated property like *schmoo*? Here it is important to note that although most concepts may be acquired via exposure to their instances, this need not always be the case. Apart from having a detector for Xs, locking can also be mediated by having a scientific theory of Xs or by deferring to the relevant experts. In the current case, getting locked to the property *schmoo* would presumably require *both* of these routes: we defer to the biologist when it comes to identifying schmoos, and the biologist (unlike Putnam's elm and beech experts, who could tell just by looking)<sup>20</sup> is in turn locked to *schmoo* via a theory of the schmoo genome. (I actually think there are problems with invoking deference in such situations; I will return to this point in chapter 4.)

So much for schmoos. But we can imagine other examples along these lines that may be more problematic. So, most of us have the concept GHOST, without (presumably) ever having encountered any ghosts. In this case, not only are there no instantiations of *ghost*, there presumably never were and never will be (ghosts being nomologically impossible entities). In this case, there plausibly isn't even any property *ghost* to get locked to.

Fodor needs a different kind of response here. In fact, he seems forced to adopt the same position that he takes for concepts such as ROUND SQUARE that are actually incoherent (rather than just nomologically impossible). That is, he has to claim that concepts for nomologically impossible entities such as GHOST are complex rather than atomic—in other words, he has to accept that informational atomism cannot handle such cases. This is problematic because, on the face of it, there are *lots* of concepts for nomologically impossible entities; Disney made a living from them.

---

<sup>20</sup> See Putnam (1975). See also Margolis (1998) for a discussion of the role of theories and deference in locking.

Doing the conceptual analysis for such cases would no doubt be fun; but is there really any reason to believe that we would have more success than with BACHELOR and KNOWLEDGE? In chapter 4 we will explore these matters in more detail, and suggest how it might be possible to avoid claiming that all such concepts are complex.

### **1.3.3. Acquisition**

One problem with Fodor's retreat from radical concept nativism is that it's not, *prima facie*, terribly clear that "locking" is all that different from his earlier notion of "triggering". It's therefore not terribly clear that Fodor's new position is less radically nativist than his old position. (This is a point that has been raised by Laurence & Margolis 1999b.) So let us examine the notions of triggering and locking in more detail.

Fodor (1981: 273) explains triggering as follows:

To a first approximation, a nativist says that the mechanism underlying the acquisition of *all* lexical concepts is brute-causal. Or, to put the same claim slightly otherwise, whereas the empiricist says that many lexical concepts are logical constructs out of primitive concepts which are, in turn, made available by the activation of the sensorium, the nativist says that the triggering of the sensorium is, normally, causally necessary and sufficient for the availability of all concepts except those that are patently phrasal.

The story about triggering might go something like this. Exposure to certain stimuli of the right kind is sufficient to make lexical concepts available. Exactly which stimuli count as being "of the right kind" is taken to be specified innately; the idea is that exposure to instances of walruses, say, makes available the (pre-existing, innate) concept WALRUS, and it does so because of the innate structure of the mind. Similarly, exposure to instances of doorknobs makes available the (pre-existing, innate) concept DOORKNOB. But while exposure to instances is necessary to make the



concept available, such experiences are not necessary to give the concept its content: the argument for innate *concepts* is an argument for innate *contents*.<sup>21</sup> This is almost identical to the empiricist story about sensory concepts, except that it applies to both sensory and non-sensory concepts. In either case, a triggering experience is necessary in order to have access to the concept. Triggering is seen as a non-rational, brute-causal mechanism which is therefore not a species of learning (cf. Samuels 2002, 2004 for some discussion of innateness and learning).

This account raises a problem. If WALRUS (say) is an innate, pre-existing concept, why should it be experiences of walruses, rather than other kinds of experience, that trigger the concept? Why should the mind be set up in this particular way, rather than allowing more arbitrary experiences to trigger particular concepts? Fodor (1981) tries to make a virtue of this problem, by pointing out that some triggers are indeed somewhat arbitrary: in the case of ducklings, to take a well-known example, almost any animate object serves to trigger MOTHER. But this is not the general case, at least for humans. It is not usually the case that we acquire concepts as a result of arbitrary experiences. It is nearly always *very specific* experiences that lead to acquisition of a concept, in particular, experiences *of entities that fall under the concept*, and even more specifically, experiences *of stereotypical instances of such entities*. The mechanism of triggering provides us with no account of why this should be so.

---

<sup>21</sup> You may be wondering how, if content is constituted by a certain kind of mind–world relation, it could be innately specified. Note that Fodor wasn't an informational atomist in 1981, so the question didn't arise, of course (back then he was an internalist, and so whatever determined content was taken to be inside the head). This being said, an informational atomist would presumably have to say something like the following about triggering. Suppose that we have an innate mechanism which underwrites the mind–world link, such as a detector for the entities in the extension of the concept. Then what is triggered is the detector, and this detector serves to establish the mind–world link. Of course, if we have an innate detector, the question arises why we need triggering at all (Bach 2000: 631 has a related query). The claim that innate concepts need to be triggered can be seen as the claim that prior to our first exposure to Xs we can't think about Xs as such, even if we have an innate mechanism for detecting them. Cf. Cowie (1999: Chapter 4) and Fodor (2001a: §3.1).

Fodor (1998a) refers to this as the doorknob/DOORKNOB problem (see §1.2.6 above; see also Cowie 1999 and Fodor 2001a).<sup>22</sup>

Now that we are more clear on the notion of triggering, we are in a better position to answer our original question: How is locking different from triggering? The general idea is that whereas triggering is *an event* that merely makes available (or “releases”<sup>23</sup>) a pre-existing contentful concept, locking is *a process* of getting oneself into a relation that gives content to an existing item of Mentalese that previously had no content (that is, that was only a mode of presentation—something like a label). On a locking account, only the mechanisms that occasion locking need to be innate. There is no need to posit innate content. Fodor doesn’t go into any detail about what these mechanisms are, or how they lead to locking. He merely offers a way for the atomist to avoid radical concept nativism: if there are such mechanisms (and a plausible story about the acquisition of appearance concepts such as RED suggests that there are), this provides an account of how we can acquire concepts that is nativist about mechanisms but not nativist about concepts.

Of course, Fodor’s account would be more convincing if he had a detailed and plausible story to tell about the psychology underlying concept acquisition: what the mechanisms are and how exactly they work. He only provides a brief sketch—after all, for his purposes he only needs to demonstrate the existence of some such mechanisms, not an account of how they work. Nevertheless, a number of authors have criticised his theory because of the lack of details on this point.<sup>24</sup> One of the aims of this thesis—in particular in chapter 5—will be to look more closely at how the process of concept acquisition that Fodor proposes might work in practice, and examine to what extent this process is a psychological one. By drawing on contemporary work in ethology on ‘learning instincts’, we will see that concept

---

<sup>22</sup> Note that if your story about innate concepts is one of innate detectors, then the DOORKNOB/doorknob problem doesn’t arise. See footnote 21 above, and §5.2 below.

<sup>23</sup> A term from ethology sometimes used by Fodor.

<sup>24</sup> For example, Cowie (1999), Landau (2000), Laurence & Margolis (1999b), Levine & Bickhard (1999) and Stainton & Viger (2000).

acquisition is driven by a complex interaction between innate constraints and environmental input, so that the traditional dichotomy between innate and learned concepts breaks down. In particular, we will look in detail at how perceptual detectors might be acquired.

In this section, we have looked at three issues concerning informational atomism—some general, some more narrow. These will provide the starting points for the next four chapters. I hope that what will develop in the course of exploring these issues is a broad and consistent picture of how, by loosening the ‘informational’ part of informational atomism in a limited and principled way, we can in the end strengthen the account of concepts provided by informational atomism.

## 2. Analyticity and meaning postulates

### 2.1. The formal status of meaning postulates

Definition theory proposed that many words were definable, and that in most cases the mental representations for these definable words had the same structure as the corresponding definition. The ultimate aim was to identify a relatively small stock of primitives out of which all lexical concepts were constructed. These primitives were innate, hence universal, and on empiricist accounts exclusively sensory (with the exception of a few logico-mathematical primitives).

Definition theory was well-suited to a formal semantic treatment, and a number of formal semanticists have developed frameworks for lexical semantics based on decompositional analyses.<sup>1</sup> One of the most influential has been Dowty (1979). On a decompositional approach, providing a semantic analysis of a definable word is straightforward. For example, a (model-theoretic) semantic analysis of the word “bachelor” might be given by a translation rule such as (1), corresponding to the definition of “bachelor” as “unmarried man”:

$$(1) \text{ bachelor}' = \lambda x [-\text{married}'(x) \wedge \text{man}'(x)]^2$$

The words “married” and “man” in the definiens could presumably be decomposed further, and would themselves appear as the definienda in other translation rules. Or consider possible translation rules for the words “dead”, “die” and “kill”:

$$(2) \text{ dead}' = \lambda y \text{---alive}'(y)$$

---

<sup>1</sup> Formal semantic treatments did not usually adopt the additional empiricist constraint that primitives should be sensory. In fact, such approaches tended to be explicitly non-psychologistic, and formal semanticists did not generally view their decompositions as having the same status as those proposed in linguistic semantics (see Dowty 1979: §4.1).

<sup>2</sup> A detailed knowledge of model-theoretic notation will not be required for what follows. The expression  $\lambda x[\psi]$ , where  $\psi$  is a well-formed formula of predicate calculus, can be read as ‘the property of being an  $x$  such that  $\psi$ ’. A prime following a natural language word indicates the translation of that word into the metalanguage of the model.

(3)  $\text{die}' = \lambda y \text{ BECOME } (\neg\text{alive}'(y))$

(4)  $\text{kill}' = \lambda y \lambda x [\text{CAUSE } (x, \text{BECOME } (\neg\text{alive}'(y)))]$

These correspond to the definitions of “dead” as “not alive”, “die” as “become dead” and “kill” as “cause to die”. Note that as well as defining words in terms of other words and logical connectives, these rules also introduce the primitives CAUSE and BECOME.

The entailment relations between lexical items can be derived from translation rules such as (1)–(4) using normal deductive methods. These rules commit the theorist to a number of claims. Most obvious of these claims is that “dead” is properly defined as “not alive”, “die” as “become dead” and “kill” as “cause to die”. But they also embody a number of other claims. For example, (2) commits us to the claim that “alive” is more basic than “dead”, since “dead” is defined in terms of “alive”. And (3) and (4) commit us to the claim that CAUSE and BECOME are primitives (see below; note that the possibility is usually left open that these primitives may differ in meaning from the corresponding English words).

All of these claims are contentious, and all have been challenged on various grounds. Take the claims regarding the definability of “dead”, “die” and “kill”. The general difficulty of defending definitions is well-known, and there are good reasons to suppose that few words can be defined (see §§1.2.2–1.2.3 above). Specific concerns have been raised in the literature concerning these particular definitions (see, for example, Fodor 1970). The other claims are also difficult to support. For the decompositional program to succeed, one member of a pair of antonyms (alive/dead, female/male, good/bad, dark/light, and so on) must be defined in terms of, and hence be more basic than, the other. In fact, though, there are not always principled grounds on which such a decision can be made.<sup>3</sup> There are also great difficulties

---

<sup>3</sup> A number of experiments by Herb Clark and others have demonstrated a so-called “markedness effect”: for certain antonym pairs, the unmarked member is retrieved or comprehended more quickly than the marked member. These results are robust, and could be regarded as supporting the claim that one member of an antonym pair is defined in terms of the other. However, these results only apply to

associated with introducing operators such as CAUSE and BECOME. These are primitives, in the sense that they are not themselves decomposable into more basic elements. But their relationship to the words “cause” and “become” is unclear.<sup>4</sup> As mentioned above, it is usually allowed that CAUSE may differ in meaning from “cause”, in order to be able to account for the fact that CAUSE appears to express only direct causation, whereas “cause” may express direct or indirect causation (similar issues arise with other such operators). To take a concrete example, although it has been proposed that “kill” has the definition “cause to die” (as expressed by the translation rule in (4)) the meanings of “kill” and “cause to die” seem to pull apart. Thus, while every case of killing counts as a case of causing to die, not every case of causing to die counts as a case of killing. For suppose that Josef is the designer of a new type of ice-pick, which fails because of a design flaw at a crucial moment on a mountain climb, sending Leon plummeting to his death. Then although Josef can be said to have caused Leon to die (albeit indirectly), he has not thereby killed him. Saying that CAUSE, unlike “cause”, expresses only *direct* causation may help with putative counter-examples such as this, but it is not after all so clear that it does. One problem is that we are not provided with any independent way of determining what counts as “direct causation”.

Another way of expressing the semantic information given by translation rules such as (1)–(4) is to use “axioms” or “laws” governing the use of words. Laws of

---

true unmarked–marked pairs as defined by linguistic criteria (such as *good–bad*, *long–short*, *fast–slow*, *intelligent–stupid*) and not to all pairs of antonymous adjectives (for example, *dark–light* is not an unmarked–marked pair by such criteria, and no markedness effect is found for this pair). Furthermore, marked adjectives do not behave psychologically like negatives. See Clark (1969a, 1969b, 1971) and Jones (1970).

<sup>4</sup> To say that CAUSE and BECOME are primitives on such accounts is not to say that they are uninterpreted. For example, the interpretation of BECOME might be given by a rule such as (i):

- (i) “BECOME( $\varphi$ )” is true at instant  $i$  iff  $\varphi$  is true at an  $i'$  that immediately follows  $i$  and is false at an  $i''$  that immediately precedes  $i$ .

(See Chierchia & McConnell-Ginet 2000: 439, from which this example was taken.) The point to note is that these kind of rules provide truth-conditional interpretations, not decompositions. Such rules are certainly not to be seen as making any psychological claims.

this sort were referred to by Carnap (1952, 1956) as “meaning postulates”. The translation rules in (1)–(4) could be recast as the meaning postulates in (5)–(8):

- (5)  $\Box \forall x [\text{bachelor}'(x) \leftrightarrow (\neg \text{married}'(x) \wedge \text{man}'(x))]$   
 (6)  $\Box \forall y [\text{dead}'(y) \leftrightarrow \neg \text{alive}'(y)]$   
 (7)  $\Box \forall y [\text{die}'(y) \leftrightarrow \text{BECOME}(\neg \text{alive}'(y))]$   
 (8)  $\Box \forall y \forall x [\text{kill}'(x, y) \leftrightarrow \text{CAUSE}(x, \text{BECOME}(\neg \text{alive}'(y)))]$

Note the use of the necessity operator, “ $\Box$ ”. Meaning postulates formulated in this way express logical necessities, and are to be viewed as placing constraints on what can be an admissible model of the natural language being studied. Since a logical necessity must be true in every model, those models in which the postulate would be false (for example, models in which “some bachelors are married” comes out as true) are eliminated. If the necessity operator were omitted, the meaning postulates would only express contingent truths rather than logical necessities.<sup>5</sup> The meaning postulates in (5)–(8) are formally equivalent to the translation rules in (1)–(4), which can be straightforwardly derived from them.

It is not the case, however, that meaning postulates and decompositional translation rules are always formally equivalent. Importantly, meaning postulates also allow for partial analyses to be given. For example, consider the meaning postulate in (9):

- (9)  $\Box \forall y \forall x [\text{kill}'(x, y) \supset \text{CAUSE}(x, \text{BECOME}(\neg \text{alive}'(y)))]$

This is identical to the meaning postulate given in (8), except that the biconditional has been replaced by a conditional. This allows us to capture the intuition that for  $x$  to kill  $y$  entails that  $x$  causes  $y$  to die, but without also committing us to the claim that  $x$  causing  $y$  to die entails that  $x$  killed  $y$ .

---

<sup>5</sup> This may be a virtue, of course, if you want to avoid making the claim that such relations are analytic. See below, as well as J. D. Fodor et al. (1975: 524); Lyons (1977: 789).

This flexibility that meaning postulates allow is extremely important. As Dowty (1979: 203) points out, the approach enables us to capture entailment relations without the need to commit ourselves to doubtful decompositional analyses. And as Jerry Fodor and others have pointed out,<sup>6</sup> it also provides a solution to what has been called the “residuum problem”—that is, cases where there is a strong intuition of semantic relatedness for which it seems impossible to give any decompositional account. Compare (10) and (11):<sup>7</sup>

- (10) a.  $\text{bachelor}(x) \Rightarrow \text{man}(x)$   
       b.  $\text{bachelor}(x) = \neg\text{married}(x) \wedge \text{man}(x)$
- (11) a.  $\text{red}(x) \Rightarrow \text{colour}(x)$   
       b.  $\text{red}(x) = \text{colour}(x) \wedge P(x)$

The entailment relation in (10a) is accounted for if we assume the decomposition in (10b). In the same way, one would expect that the entailment relation in (11a) is to be accounted for by some decomposition such as (11b). But this raises the question of what the predicate *P* could possibly stand for. What predicate has a meaning such that when combined with “coloured” it gives “red” (other than “red” itself)? There does not seem to be a possible meaning for such a predicate. This problem is apparently not limited to sensory concepts, either; similar difficulties arise in explaining the entailment relations between “dog” and “animal”, say, or between “chair” and “furniture”.

---

<sup>6</sup> See Fodor (1975: 149 ff., 1998a: 108 ff.), J. D. Fodor (1977: 150 ff.) and J. D. Fodor et al. (1975: 527).

<sup>7</sup> This seems like a good moment to introduce some conventions. I adopt the standard convention of using “ $\supset$ ” for material implication, “ $\Rightarrow$ ” for entailment and “ $\vdash$ ” for deducibility (when an entailment is deducible in a system, it is derivable using the deductive rules available to that system; this is what Sperber & Wilson 1995: 84 refer to as “logical implication”). Note, of course, that applying a necessity operator to a material implication, as in (9), results in an entailment. The symbol “ $\rightarrow$ ” is reserved for denoting “rewrites as” in substitution rules (discussed below). Occasionally I will quote expressions that include variables; I will not always bother with corner quotes.



Meaning postulates can provide a solution to this problem, since they allow entailments to be captured without any commitment to a decompositional analysis, as in (12).

(12)  $\Box \forall x [\text{red}'(x) \supset \text{colour}'(x)]$

This amounts to merely a statement of the entailment, rather than an explanation of it, of course. As Quine pointed out to Carnap,<sup>8</sup> it follows that such meaning postulates cannot play a role in, for example, deciding where to draw an analytic–synthetic distinction. We will return to this issue below. First, I want to consider what psychological role meaning postulates could play.

## 2.2. The psychological status of meaning postulates

Formal semantic analyses do not normally aim at psychological reality. Providing a formal analysis of (some fragment of) a language is an important goal in itself.<sup>9</sup> But since we are engaged here in a naturalistic enterprise, it is important to consider the psychological status of the claim that intuitions of semantic relatedness or entailment relations are to be captured by meaning postulates.

The first thing to note is that from a psychological perspective axioms might not be the best way of thinking about meaning postulates, because axioms are by nature representational rather than computational. As Carroll (1895) highlighted, a logical system cannot consist of axioms with no inference rules, although such a system could consist of inference rules and no axioms (this point is underlined by Sperber & Wilson 1995: Chapter 2; see also Millikan 1993: §7). The way that Fodor talked about meaning postulates suggests that he, too, viewed them not as axioms but as inference rules. For example, he said “...if we want  $F$  to entail  $G$ ...we should add  $\lceil F \rightarrow G \rceil$  to the inference rules” (1975: 149) and “...the representation of ‘John is a bachelor’ does determine the entailment ‘John is unmarried’ if...the inference rules

---

<sup>8</sup> See Quine (1953a).

<sup>9</sup> For a discussion see Dowty (1979: Chapter 8).

which apply to that representation include *bachelor*  $\rightarrow$  *unmarried*” (1975: 150). He also explicitly endorses Carroll’s point in Fodor (1983: 9): “...knowledge doesn’t eventuate in behavior in virtue of its propositional content alone. It seems obvious that you need mechanisms to put what you know into action...”.<sup>10</sup>

Consider the entailment ‘ $\text{dog}(x) \Rightarrow \text{animal}(x)$ ’, and suppose that we have this entailment mentally represented as a meaning postulate. We are suggesting two ways that this meaning postulate could be represented: either as an axiom as in (13), or as an inference rule as in (14):

(13)  $\Box \forall x [\text{DOG}(x) \supset \text{ANIMAL}(x)]$

(14)  $\varphi \text{DOG } \psi \rightarrow \varphi \text{ANIMAL } \psi$ <sup>11</sup>

The axiom in (13) is just a statement of fact (albeit marked as a necessary truth), and as such we make the assumption that it is stored with other propositions in our encyclopaedic knowledge. As an inference rule, we make the assumption that (14) is part of our computational deductive system (our mental logic—or what I shall refer to, following Sperber & Wilson 1995, as our “deductive device”). In addition to the format of storage being different (representational vs. computational), a number of other differences are also implied, such as the reliability of application and speed of operation.

Consider the following. An inference rule that forms part of our mental logic will presumably apply to any representation that meets its input conditions (that is, any representation of the form stated on the left of the rule), subject to some important restrictions. First, the representation must be entertained in the right way in order to

---

<sup>10</sup> But note that Fodor (1975: 150) also spoke of meaning postulates being stored in long-term memory. J. D. Fodor (1977: 153) also refers to meaning postulates as inference rules. See Cormack (1998: 39) for further discussion.

<sup>11</sup> This is approximately the format used in Sperber & Wilson (1995). Inference rules of this kind take the form of substitution rules, where ‘ $\rightarrow$ ’ denotes ‘rewrites as’;  $\varphi, \psi$  denote strings. Some restrictions must be placed on the set of strings over which  $\varphi, \psi$  can range, as discussed in Cormack (1998: §1.2). Depending on the precise nature of the restrictions, we may need ANIMAL OF A CERTAIN KIND rather than simply ANIMAL in the substitution rule.

trigger the application of the rule. I will not go into great detail as to what being “entertained in the right way” amounts to, since the specifics are not important for our present concerns,<sup>12</sup> but we can think of it as something like being represented by our current thought processes—that is, as appearing in the buffer of our deductive device (rather than just being represented in long-term memory, say). What we want to avoid is the claim that such rules apply to any suitable representations in our encyclopaedic knowledge, whether they are currently represented in thought or not, as this would lead to a computational explosion. Or, as Harman (2002: 174) puts it, there is no reason to “clutter one’s mind with trivialities just because they follow from other things one believes”. The second restriction is needed because we must allow that the application of an inference rule can be pre-empted by the application of other rules or procedures. The considerations in Sperber (2005) make this clear. Sperber argues that inference rules or procedures (even if they are currently represented in thought) do not apply mandatorily, but must compete for cognitive resources (as all computations must). We can imagine, Sperber suggests, that a frog’s fly-catching reflex may be pre-empted by an escape reflex in the presence of both a fly and a predator; this would certainly make sense. In the case of humans, who of course have a far more complex psychology less reliant on simple reflexes, we can imagine a more intricate (perhaps relevance-based) competition for resources between procedures. Again, what must be avoided is a computational explosion, which could occur if we assumed that all inference rules associated with the concepts represented by our current thought processes applied automatically or mandatorily. As Sperber notes, it may be that the circumstances in which one procedure pre-empts another are quite rare, in which case a procedure may appear to apply mandatorily. Sperber’s position seems to be confirmed by the phenomenon of inattentional

---

<sup>12</sup> For a detailed discussion, see Sperber & Wilson (1995: Chapter 2).

blindness, where (usually in contrived experimental conditions) subjects can fail to be aware of events that would normally be highly salient.<sup>13</sup>

I do not wish to make too much of these caveats, however. The present point is that an inference rule will, *ceteris paribus*, reliably apply if its input conditions are met. I wish to contrast this with the situation for axioms such as (13) stored as encyclopaedic knowledge. These could not reliably apply, even *ceteris paribus*. For in order to be used by our deductive device, an axiom such as (13) would have to appear in the buffer of the device together with a proposition containing a relevant concept (DOG, as it might be), so that a suitable inference rule could apply (presumably *modus ponendo ponens*, given that the axiom contains a material implication). That is, an axiom is just another representation, which must be retrieved from memory and appear along with other suitable representations in our deductive device in order for the appropriate inference rule to apply and derive a conclusion. But it is clear that we do not have infallible access to our encyclopaedic knowledge, and it cannot be expected that, given well-known constraints on memory retrieval, axioms will always or even reliably be retrieved along with propositions containing a relevant concept. If we want to ensure priority application of a meaning postulate, it had better be represented as an inference rule such as in (14).

Considerations of speed of application point in the same direction. Spontaneous inference is fast, whereas accessing stored encyclopaedic information is a relatively slow search process.<sup>14</sup> It therefore seems unlikely that spontaneous inference relies on axioms stored as encyclopaedic information, since we do not have suitably quick access to the totality of such information. These considerations militate against an account of meaning postulates as axioms. Indeed, it is presumably because deriving certain information reliably and quickly is of importance to the organism that certain

---

<sup>13</sup> Such as a gorilla walking through a basketball game in the middle of play. See Simons & Chabris (1999). For a detailed account of inattention blindness, see Mack & Rock (1998), and Mack (2003) for a recent overview.

<sup>14</sup> Shastri & Ajjanagadde (1993) give an estimate of as little as a few hundred milliseconds for spontaneous inference.

knowledge has been formalised as procedures (that is, meaning postulates), rather than just being stored in the encyclopaedia.

If it is true that we have meaning postulates (in the form of inference rules) in addition to encyclopaedic knowledge, then the possibility of conflicts between the two obviously arises. Evidence of such conflicts—that is, evidence that spontaneous inference and reflective thought can produce different results—would indeed be evidence for distinct mechanisms underlying these thought processes. And such a situation does in fact arise quite commonly. Convincing fakes may continue to trigger inappropriate inferences even after we have discovered that they are fakes; we need to consciously override this with reflective thought (think of computer-generated dinosaurs in the movie “Jurassic Park”, or some of the more realistic toy animals). Similarly, even though we may know that whales are mammals, they may continue to strike us as fish. In such cases, spontaneous inference conflicts with encyclopaedic knowledge.<sup>15</sup>

In this thesis, I will follow Sperber & Wilson (1995) in arguing that, although we may have any number of necessities such as (13) represented in our encyclopaedic knowledge, we also have meaning postulates in the form of inference rules such as (14) available to our deductive device. These are the logical entries of concepts, and they support our spontaneous inferential capacities. Now, it may seem somewhat strange to consider that we have inference rules for non-logical concepts such as DOG. But in fact there seems to be no principled reason for making a distinction between the logical and non-logical vocabularies in this regard, and like Sperber & Wilson I will argue below in favour of earlier proposals of Fodor and others<sup>16</sup> that no such logical/non-logical distinction should be made. Henceforth, when I refer to meaning postulates, it is in the sense of mentally-represented inference rules of the kind in (14).

---

<sup>15</sup> For more detailed discussion of such conflicts, see §4.3.1.

<sup>16</sup> Fodor et al. (1980: 269 ff), Katz (1972).

The case that I will make is this. It is widely (though not universally) accepted that meaning postulates are necessary to account for the content of logical terms. I will argue that no principled distinction can be made between logical and non-logical terms, and therefore that we should not rule out the possibility that meaning postulates are also associated with non-logical terms. I will further argue that this does not require us to adopt a purely inferential role semantics—rather, meaning postulates can be combined with something like the account of non-logical content that is proposed by informational atomism. In essence then, the distinction that I am suggesting is between logical and non-logical *content* (the former to be captured via meaning postulates, the latter via atomism)<sup>17</sup> rather than between a logical and non-logical *vocabulary*.

I will proceed as follows. In the next section I argue that no principled distinction can be made between the logical and non-logical vocabularies. This is problematic for those atomists who wish to combine an inferential role account for the logical vocabulary with an atomistic account for the non-logical vocabulary. Next, I will sketch Fodor's arguments against inferential role accounts of conceptual content in general, which I endorse. This would seem to leave only one option: a fully atomistic account of content, including for the logical vocabulary, as recently proposed by Fodor (2004a, 2004b). I argue, however, that the possibility outlined above—combining meaning postulates with an atomistic account of non-logical content—is still open, and is to be preferred over a fully atomistic account.

---

<sup>17</sup> I will leave open the question of whether these two categories are exhaustive, however. For example, there are aspects of meaning that are non-truth-conditional but also non-referential, and which are perhaps best captured by inferential/heuristic procedures that would not count as meaning postulates.

### 2.3. The logical and non-logical vocabularies

Fodor et al. (1980) note that logic is commonly thought of as providing a reconstruction of our intuitions about the validity of arguments such as that in (15), but not of arguments such as that in (16).

(15) John left and Mary wept, therefore Mary wept

(16) John is a bachelor, therefore John is unmarried

(Fodor et al. 1980: 269)

The traditional view in logic has been that this is because the argument in (15) is valid in virtue of the meaning of ‘and’, which is part of the logical vocabulary, whereas the argument in (16) is valid in virtue of the meaning of ‘bachelor’, which is part of the non-logical vocabulary. But why, given that we have the intuition that both (15) and (16) are valid arguments, should we draw this kind of distinction?

In discussions of the logical and non-logical vocabularies there is some variation (and vagueness) in which are to be considered the logical terms, as can be seen from the following examples:

(17) ...such expressions as “all, some, not, or, equals, if then”, and “and”.

(Fodor et al. 1980: 269)

...“+”, “and”, “all” and the like...

(Fodor 1990b: 110)

...the terms usually called “logical” (that is, the usual connectives and quantifiers plus identity)...

(Hanson 1997: 375)

...these terms, the ‘logical constants’, include the usual ‘and’, ‘or’, ‘not’, ‘for all’, ‘exists’, perhaps ‘=’, and terms definable in terms of these.

(Blanchette 2001: 132)

This is perhaps not surprising. After all, it's not clear that there is any principled way of determining which words are contained in these respective vocabularies. The logical vocabulary is made up of those words that are introduced as logical terms in the particular logical system(s) one is employing; all other words are considered to be part of the non-logical vocabulary. But this does not help us very much, for the various systems of logic have been developed in part to explain our pre-theoretic intuitions of logical validity and do not themselves provide a principled way to answer the question "how can we determine if a given word is part of the logical or non-logical vocabulary?". Of course, general considerations of simplicity—Ockham's razor—apply to logic as much as other fields, so that one system of logic may be regarded as giving a superior account of certain phenomena to another. But these considerations only apply once we have made the prior decision as to what facts we want our logical system to explain. For example, if we decide that our logic should explain the validity of arguments that turn on the meaning of modals, then we can try to develop a modal logic that provides an interpretation of modal terms, so that modal terms would be part of the logical vocabulary. If not, then our logic will be non-modal and modal terms will be considered part of the non-logical vocabulary. The same is true of tense, aspect and other linguistic phenomena. Such decisions are ultimately a matter of stipulation.<sup>18</sup> The view we are considering here is that logical accounts of natural language phenomena should seek to explain the totality of our pre-theoretic intuitions of validity, including of arguments such as (16)—that is, we should regard all lexical items as potentially having logical properties. This is not a new idea. The rejection of a clear distinction between logical and non-logical terms was something that Tarski suggested in 1936:

Underlying our whole construction...is the division of all terms of the language discussed into logical and extra-logical. This division is certainly not quite arbitrary. If, for example, we were to include

---

<sup>18</sup> But see Tarski (1966/1986) and Sher (1991, 1996) for attempts to develop a principled way to distinguish logical from non-logical terms.



among the extra-logical signs the implication sign, or the universal quantifier, then our definition of the concept of consequence would lead to results which obviously contradict ordinary usage. On the other hand, no objective grounds are known to me which permit us to draw a sharp boundary between the two groups of terms. It seems to me possible to include among logical terms some which are usually regarded by logicians as extra-logical without running into consequences which stand in sharp contrast to ordinary usage. In the extreme case we could regard all terms of the language as logical. (Tarski 1936/1956: 418–419)

More recently, a number of authors have argued for the same view (see, for example, Cresswell 1978, Dowty 1979: §4.1, Hanson 1997 and Varzi 2002).

Nevertheless, the temptation to maintain a clear line between the logical and non-logical vocabularies is a strong one, and a number of theorists have succumbed to it for various reasons. For example, some atomists (most notably, Jerry Fodor prior to his millennial change of heart) reject inferential role accounts for the content of the non-logical vocabulary, but nevertheless accept that such accounts may be needed for the content of logical terms. The reason is that, *prima facie*, the meaning of words such as ‘and’ *just is* their inferential role.<sup>19</sup>

The fact that such atomists have generally been unable to devise any plausible alternative account for the content of the logical vocabulary has meant that they must rely on distinguishing between logical and non-logical words. That is, they accept that inference rules are needed to account for the meaning of the logical vocabulary, but they maintain that the same is not true of the non-logical vocabulary (see Fodor 1987: 78; 1994: Chapter 3). Such a position of course requires that a clear distinction

---

<sup>19</sup> Clearly, there are other quite obvious *possible* ways of accounting for the meaning of logical words such as ‘and’. For example, truth tables could be used instead of inference rules. But bear in mind that the issue at hand is psychological, not logical, and I take it that the explanation for such cases is not that we have truth tables in our heads. (Not everyone takes it this way. Mental models theorists such as Johnson-Laird do claim that people use something like truth tables, rather than logical rules, in reasoning. See, for example, Johnson-Laird 1993.)

can be drawn between the logical and non-logical vocabularies. For example, Fodor has in the past stated this explicitly:

Presumably a language that didn't have 'not' couldn't have 'if', and maybe a language that didn't have sentence conjunction couldn't have predicate conjunction. But there is no reason at all to suppose that the logico-syntactic vocabulary is itself interdefined with the *non*-logical- vocabulary. (Fodor 1994: 76, original emphasis)

These atomists are generally not explicit about *which* words they take to make up the logical vocabulary (typically, they seem to have in mind the standard connectives and quantifiers), nor are they clear regarding what grounds are to be used to make the distinction.<sup>20</sup> Such a position raises a number of questions. What would be the status of non-standard quantifiers such as 'many' or 'most'? How about modals, tense and aspect? Other words also support entailments. For example, there is an entailment relation between 'either'/'both' and 'two'. How are we to account for this? This in turn raises the question of whether (some) numerals are included in the logical vocabulary. As discussed above, in the end what counts as a logical term is a matter of stipulation.

So it looks as though those who wish to draw a logical/non-logical distinction just have to stipulate which terms are logical and which are not. But in fact things are worse than this. An additional problem is that there are all sorts of words which appear to have properties of both the logical and non-logical vocabularies. And if there are indeed words that share some properties of logical terms, but which also have non-logical content, then this suggests that no logical/non-logical distinction can in fact be drawn. Consider the example of 'inside' (cf. Sperber & Wilson 1995: 105).

---

<sup>20</sup> Fodor had suggested (personal communication, 2001) that no logical term can appear as the constituent of a closed, atomic sentence.

- (18) a.  $x$  is inside  $y$   
       b.  $y$  is inside  $z$   
       —————  
 (19)  $x$  is inside  $z$

(18)–(19) is an example of a valid inference. The relation ‘is inside’ has certain formal properties: it is *transitive* (as just exemplified), *irreflexive* (nothing is inside itself), and *asymmetric* (if  $x$  is inside  $y$ , then it is never the case that  $y$  is also inside  $x$ ). These properties may be captured by a set of inference rules in the same way that the properties of a logical connective such as ‘or’ are captured. And yet ‘inside’ cannot simply be added to the logical vocabulary; at least part of its meaning derives from its non-logical content. The point about the logical vocabulary is that its content *just is* its inferential role. But the content of ‘inside’ *cannot* just be its inferential role; if it were, then it would be synonymous with ‘below’, which has identical logical properties (like ‘inside’, ‘below’ is transitive, irreflexive and asymmetric).<sup>21</sup>

Now, it may be argued that the formal properties of ‘inside’ are not semantic at all. They might rather be taken as metaphysical properties of containment, and our ability to make inferences along the lines of (18)–(19) might have nothing to do with rules of inference (perhaps such spatial reasoning involves the construction of some kind of mental model, for example). So be it. The point I wish to make is only that some words have logical properties in addition to their non-logical content, and that in some cases we may wish to capture these logical properties in the same way that we capture the properties of the standard logical terms, through inference rules. It is, of course, an empirical issue which rules of inference our mental logic actually employs.

Less contentious, perhaps, are examples of words which have referential/descriptive content, but which also exhibit the properties of one or other of the

---

<sup>21</sup> Conversely, it cannot be the case that the logical properties of ‘inside’ are general to spatial prepositions, since not all spatial prepositions have the same logical properties (‘beside’, for example, is non-transitive, irreflexive, and symmetric).

standard logical connectives. For example, adjectives such as 'raw' and 'dark' share the logical properties of the operator 'not', for 'raw' entails 'NOT cooked' and 'dark' entails 'NOT light'. Connectives such as 'since' and 'because' share the logical properties of 'and', for 'x since/because y' entails 'x AND y'. The word 'sibling' shares the logical properties of 'or', for 'sibling' entails 'brother OR sister'. Similarly with 'unless': 'x unless y' entails 'x OR y'. In none of these cases do these logical properties exhaust the meaning of the word, so it would not be possible to claim that these are logical words. Presumably, however, their logical properties must be accounted for in the same way as with the corresponding logical constant—via inference rules. Note that in these cases, the question is no longer whether certain logical properties are to be captured by inference rules. Since the logical properties in question are the same as those exhibited by one of the standard logical connectives, we would already be committed to the claim that we have such rules in our mental logic. It would therefore seem natural to account for these logical properties by making use of the same rules.

A similar point can be made with regard to non-truth-conditional connectives such as 'but' and 'although'. Are these logical or non-logical words? Most people agree that 'but' is logically equivalent to 'and', so that it would be governed by the same inference rules. Clearly 'and' and 'but' do not have the same meaning, however—in addition to its truth-conditional meaning, 'but' introduces some non-truth-conditional notion of 'denial of expectation' or 'contrast' (see Gazdar & Pullum 1976: 223). Here is another case where the logical and non-logical vocabularies appear to be interdefined.

So it is not only that any logical/non-logical distinction is unmotivated. It would seem to be impossible in practice to draw such a distinction. However, there are well-known difficulties in defending such a view. For if we accept that the content of logical words is given by their inferential role, and if we also deny that any division can be made between logical and non-logical words, then it seems that we are committed to claiming that the content of all words is given by their inferential role.

This is a view known as ‘inferential role semantics’, and is precisely the view that atomistic theories of content have set themselves up in opposition to.

Where we have got to is this. Inferential role theorists do not need to make a principled distinction between the logical and non-logical vocabularies. They extend the widely-accepted inferential role account of the meaning of logical connectives to the non-logical vocabulary as well. Atomists argue against inferential role semantics, for reasons which are examined in some detail below. Some atomists, however, do accept an inferential role account in the limited case of the logical vocabulary. This requires that a principled distinction can be made between logical and non-logical terms, something I have argued against above. The other possibility open to atomists would be to reject inferential role accounts completely, and adopt an atomistic account of the logical vocabulary as well. This is a position that Jerry Fodor has recently advocated (see Fodor 2004a, 2004b), and is discussed in more detail in §2.5 below. First, we will look at the arguments against an inferential role semantics.

## **2.4. Problems with inferential role semantics**

### **2.4.1. *Content-constitutive inferences***

The inferential role of an expression is the totality of the inferences in which that expression plays a role. According to inferential role semantics, the conceptual content of an expression is determined (in some specified way) by its inferential role. Take ‘and’, for example. A competent user of ‘and’ will be disposed to accept as valid certain inferences involving ‘and’, such as:

- (20) a. Josef is mad. Leon is scared. Therefore, Josef is mad and Leon is scared.  
b. Josef is mad and Leon is scared. Therefore, Josef is mad.  
c. Josef is mad and Leon is scared. Therefore, Leon is scared.

According to inferential role semantics, ‘and’ gets its content in virtue of the fact that a certain distinguished set of inferences involving it are taken to be valid, namely, inferences of the following general form:

(21) a. *A, B. Therefore, A and B.*

b. *A and B. Therefore, A.*

c. *A and B. Therefore, B.*

In general, the inferences that are constitutive of an expression’s content may or may not be a proper subset of the valid inferences involving that expression.<sup>22</sup> This means that we need some criterion for determining, of all the valid inferences involving a given expression, which are constitutive of that expression’s content.

To put it another way, there are two possibilities open to the inferential role theorist: either to provide a way of determining *which* inferences are content constitutive or to allow that *all* inferences that an expression enters into are constitutive of its content. Fodor argues that either of these routes leads to unacceptable consequences.

#### **2.4.2. *Meaning holism***

Let us take the second possibility first: that all of the inferences in which an expression plays a role are constitutive of its content. The idea, roughly, would be that any meaningful expression (a word, say, or a concept) gets its meaning from its place in a larger system of such meaningful expressions (a language, say, or a conceptual system). If this is true, then it means that any change to the system will affect all parts of that system. So changes to my beliefs about the presence of water on Mars would in principle bring about changes, however minimal, to the content of my DOG concept. The general point is that since, according to the proposal we are considering, expressions get their meaning from their place within the larger system

---

<sup>22</sup> For a discussion of inferential role semantics for logical terms see Kalderon (2001). Prominent advocates of inferential role semantics more generally include Block (1986, 1993), Boghossian (1993, 1994), Field (1977), Lycan (1984) and Peacocke (1992).

as a whole, it follows that expressions from distinct systems cannot ever be equivalent in meaning. This is a doctrine that is known as meaning holism.<sup>23</sup>

Now, some may regard meaning holism as unacceptable in itself. Here's Fodor and Lepore:

...once you start identifying the content of a belief with its inferential role in a system of beliefs (*mutatis mutandis*, the meaning of an expression with its inferential role in a language), it's hard to avoid proceeding to identify the content of a belief with its *whole* inferential role in the system of beliefs. And, having gone that far, it's hard to avoid the conclusion that if belief systems differ at all with respect to the propositions they endorse, then they differ completely with respect to the propositions they endorse. ... This is a well-greased, and well-travelled, slippery slope; having arrived at the bottom, one finds oneself accepting such *prima facie* outlandish doctrines as that no two people ever share a belief; that there is no such relation as translation; that no two people ever mean the same thing by what they say; that no two time slices of *the same* person ever mean the same thing by what they say; that no one can ever change his mind; that no statements, or beliefs, can ever be contradicted (to say nothing of refuted); and so forth. (Fodor & Lepore 1991: 331, original emphasis)

Or, more succinctly (and without Lepore's restraining influence):

...Meaning Holism really is a *crazy* doctrine. (Fodor 1987: 60, original emphasis)

Fodor regards the considerations that he and Lepore set out as a reductio of meaning holism.<sup>24</sup> He has even formulated a non-negotiable condition on theories of concepts which rules out any theory having the relativistic consequences that meaning holism exhibits.<sup>25</sup> And for those in need of another argument against meaning holism, Fodor & Lepore (1991) offer the following. Meaning holism is

---

<sup>23</sup> For a detailed discussion, see Block (1986), Fodor & Lepore (1991, 1992), Cain (2002: 122 ff.).

<sup>24</sup> See Fodor (2004a: 35).

<sup>25</sup> See Fodor (1998a: 28 ff.) and §1.2.1 above.

incompatible with the principle of compositionality—that is, with the principle that the meaning of a sentence (thought) is composed from the meanings of its parts. Since, it is further argued, the principle of compositionality is non-negotiable, meaning holism must be rejected.

We will now take a look at this argument in more detail. But first, a disclaimer. In what follows, I'm not going to present detailed arguments in favour of the principle of compositionality; its status has not been seriously called into question, and in any case it is common ground in the debate between Fodor and inferential role theorists.<sup>26</sup> Suffice it to say that we need compositionality to explain both the productivity and systematicity of language and thought. Productivity is the ability of a system to generate a potentially infinite number of distinct expressions from a finite base. Systematicity is the property of a system such that if it can represent *P* then it can also represent variants of *P* which differ in the systematic recombination of its constituents (for example, 'John loves Mary' and 'Mary loves John'). Compositionality is needed to explain both: it's only because the content of a complex expression is composed from the contents of its parts (plus the syntax) that an indefinite number of distinct contentful expressions can be generated, and the use of a finite base to generate them explains the existence of systematically related variants.

So to Fodor and Lepore's argument for meaning holism being incompatible with compositionality. Meaning holism identifies the content of an expression with its inferential role *tout court*. For example, the inferential role of 'black cat' might

---

<sup>26</sup> See Fodor & Lepore (2002) for a detailed discussion of compositionality. Of course, acceptance of the principle of compositionality is not universal, but it is nearly so (some connectionists, though, deny it). Apart from this, there is considerable doubt as to whether natural language could have a compositional semantics (its context-dependent nature and other phenomena would appear to rule this out), but this is different from calling into question the principle of compositionality itself. If natural language doesn't itself have a compositional semantics, then it just means that thought must—since language expresses thoughts, only one of the two can have primitive (that is, underived) content, and whichever one does must respect compositionality. All the available evidence suggests that it's thought that is primitive, hence compositional (see Fodor 2001b; see also Fodor & Lepore 2005).



include not only such inferences as ‘black cat  $\rightarrow$  black’ and ‘black cat  $\rightarrow$  cat’, which compositionality requires, but also such inferences as (if one happens to believe that black cats are unlucky, say) ‘black cat  $\rightarrow$  unlucky’. The problem is that this last inference doesn’t seem to depend on the constituents of ‘black cat’, since the inference from ‘black cat’ to ‘unlucky’ doesn’t depend on an inference from ‘black’ to ‘unlucky’ or on an inference from ‘cat’ to ‘unlucky’. Rather, being unlucky is a property peculiar to black cats *qua* black cat. This implies that the inferential role of ‘black cat’ depends not only on the inferential roles of its constituents, but also on our particular beliefs about black cats. And this in turn implies that inferential roles *tout court* are not compositional. It follows that content cannot be identified with inferential roles *tout court*, and that meaning holism must therefore be rejected.

So there are a number of reasons why it is unwise for the inferential role theorist to allow that all the inferences an expression enters into are constitutive of its content. In which case, the only other option is to provide a way of determining *which* inferences are content constitutive.

### **2.4.3. Molecularism**

Among those who reject semantic holism, the traditional way to determine which inferences are content constitutive has been by employing some notion of analyticity—the claim being that all and only the analytic inferences that an expression enters into are constitutive of its content. But Fodor & Lepore (1991) point out serious difficulties with such an approach.

First, note that compositionality is respected only at the cost of apparent circularity. It’s true that the analytic inferences of an expression *just are* the ones that are warranted by the meanings (analytic inferences) of its constituents, so compositionality is straightforwardly satisfied. But one cannot fall back on analyticity to distinguish the content-constitutive inferences of an expression, since the notion of analyticity is itself explicated in terms of meaning. Rather, to avoid

circularity, one would have to employ a notion of analyticity explicated in terms of some non-semantic property.<sup>27</sup>

Even if this potential circularity can be avoided, however, there is a more fundamental difficulty. Quine (1953a) has cast serious doubt on the very notion of analyticity. If, say Fodor & Lepore, Quine's arguments are correct (as they are widely regarded to be) then there are no expressions that are true or false solely in virtue of what they mean, hence there are no analytic inferences. So, it follows that if a non-holistic inferential role semantics is to be maintained, then we are owed a refutation of Quine's arguments against the analytic/synthetic distinction. The prospects of this do not look good.

The conclusion is that either inferential role semantics (in its holistic form) leads to a violation of the fundamental principle of compositionality, or (in its molecular form) leads to a reliance on the discredited notion of an analytic/synthetic distinction. Fodor & Lepore therefore reject inferential role semantics.

Some atomists, as we have seen above, do endorse an inferential role account for the logical vocabulary. On the assumption that the logical and non-logical vocabularies are not interdefined, this gives them a possible way to avoid the arguments against molecularism. That is, they rely on the possibility of drawing a logical/non-logical distinction that does not require discredited notions of analyticity. I have argued above against the assumption that the logical and non-logical vocabularies are not interdefined. But a case can also be made that Quine's arguments against an analytic/synthetic distinction apply equally to inferential role accounts of the logical vocabulary.

As Boghossian (1997: 354) points out, if we claim that the meaning of logical words is constituted by the inferences they enter into, we must still ask *which ones*. On what basis are we to decide which of the inferences a logical word enters into are to be taken as constitutive of its content? In particular, how are we to decide between

---

<sup>27</sup> See also Boghossian (1993) for discussion of this point.

an obviously valid but non-constitutive inference, and a content-constitutive inference (or if you prefer, between an obviously true but non-defining sentence, and a sentence which implicitly defines the meaning of some logical term)? Quine's arguments are precisely that we cannot make such a decision (see, for example, Quine 1935/1976, 1954/1976).

Consider the case of disjunction. Many people are of the view that  $p \vee \neg p$  (the so-called 'law of excluded middle') could plausibly be false. Witness, for example, the range of many-valued logics that have been developed—motivated by concerns ranging from issues of free will (Łukasiewicz 1920/1970) to the peculiarities of quantum mechanics (Reichenbach 1944). Others have suggested that  $p \vee \neg p$  may be true but non-constitutive (for example, Fine 1982). These possibilities might seem somewhat controversial, but as Boghossian (personal communication, 2001) points out, this is to be expected precisely because we lack a convincing theory of how to pick out the defining sentences. We will look at this issue in more detail in chapter 3.

Furthermore, in addition to the arguments against molecularism that we considered above, Fodor has recently outlined another argument, which he takes as showing that inferential role accounts for the content of logical terms are circular.

#### **2.4.4. *The circularity argument***

Fodor (2004a) argues that inferential role accounts for the content of (in particular, logical) terms are viciously circular. His argument proceeds as follows. First, he notes the links between inferential role semantics and implicit definition.

Successful implicit definitions are supposed to provide examples of how the content of a concept might be determined by the rules of inference that apply to it, and of how compliance with such rules might be constitutive of having the concept. (Fodor 2004a: 40)

For logical terms, which have no referential content, such implicit definitions are plausibly the whole story. Implicit definitions (or possession conditions on concepts of the kind proposed by Peacocke 1992) provide introduction and elimination rules

for the concept. The claim that Fodor argues against is not that such rules can adequately define the concept, but rather the claim that grasping a concept can be identified with grasping its introduction and elimination rules (say, in the form of an implicit definition or possession condition).

For example, Peacocke (1992: 6) formulates the possession condition for the concept of conjunction by stating that “conjunction is that concept  $C$  to possess which a thinker must find transitions that are instances of the [forms in (22)] primitively compelling, and must do so because they are of these forms”.<sup>28</sup>

- (22) a.  $p, q / pCq$   
b.  $pCq / p$   
c.  $pCq / q$

Fodor’s argument rests on what Peacocke means by the use of “form” when he says that a thinker must find certain inferences compelling “because they are of these forms”. Peacocke needs such a caveat for two reasons. First, it is necessary to avoid circularity, since the most obvious reason to accept inferences of the form in (22) would be that the thinker already has the concept CONJUNCTION. But it would be clearly circular to formulate a concept’s possession condition by employing that very concept. Second, the caveat is necessary in order to rule out situations where a thinker could come to find inferences of the form in (22) compelling not by grasping conjunction, but in virtue of something else (a reflex, the result of a head injury, and so on).

Fodor suggests that there are only two obvious ways to interpret “form”: either Peacocke could have in mind *logical* form, or he could have in mind *syntactic* form. He argues that neither choice will work. For logical form preserves precisely the logical terms (such as ‘and’) that one is trying to give an account of. Finding a

---

<sup>28</sup> For a thinker to find an inference ‘primitively compelling’, according to Peacocke, is for that thinker (i) to find the inference compelling, (ii) not to do so as a result of inferring it from something else, and (iii) not to necessarily take correctness of the inference as answerable to anything else.

sentence of the form in (22), for example “If John swims and Mary swims then John and Mary swim”, compelling in virtue of its logical form would require prior possession of CONJUNCTION. Syntactic form will not work either, because notions such as validity and correctness are not syntactic. The form of inference ‘ $p$  and  $q \rightarrow p$ ’ is correct not in virtue of its syntax, but in virtue of the meaning of ‘and’, and therefore accepting such an inference shows not syntactic knowledge, but grasp of CONJUNCTION. Fodor concludes that there is no non-circular way of formulating the possession condition of a logical term such as ‘and’. This leads him to reject inferential role accounts for the content of logical terms.

## **2.5. ‘Pure’ informational atomism**

### ***2.5.1. Fodor’s new proposal***

Where does this leave Fodor? As we have seen, he has long argued for an account of the content of non-logical concepts based on informational atomism. On this view, (lexical) concepts have no internal structure, and get their content from mind–world links. However, there has so far been a consensus (which included Fodor himself) that inferential role is the only conceivable way to account for the content of logical concepts. In many ways, it has been the success of inferential role semantics in accounting for the content of logical terms that has motivated the extension of such accounts to the non-logical vocabulary. If Fodor is now rejecting inferential role accounts of the logical vocabulary, he owes us some alternative explanation.

It is not surprising what sort of account Fodor has in mind. He comments that he is “much inclined to think that Cartesians are right about what concept possession is; and that it more or less follows that conceptual content is atomistic and that meaning is reference” (2004a: 47). Now, it’s pretty clear how this might work in the case of tigers. That is, ‘tiger’ expresses TIGER which expresses the property of *tigerhood*. The content of TIGER is constituted in virtue of certain (asymmetrically dependent, counterfactual-supporting) nomological links between TIGER and tigers. An

analogous proposal for logical terms might be something like this: ‘and’ expresses AND which expresses the property of *andness*. The content of AND would then be constituted in virtue of its nomological links with (instances of) conjunction. But this is certainly not what Fodor has in mind.<sup>29</sup> The point about the logical constants such as ‘and’ is that they apparently don’t have any referential meaning. That’s why even atomistic theories of content have tended to treat them differently.

So what does Fodor’s new proposal consist in? He doesn’t provide us with many clues. To begin with, it is clear that he doesn’t object to introduction and elimination rules such as (22) as ways of stating the meaning of a logical term. To put it another way, he has no problem with implicit definitions for ‘and’, say, *qua* definition (in the same way as he doesn’t object to the fact that ‘bachelor’ can be defined as ‘unmarried man’). What he disagrees with is the claim that these introduction and elimination rules (/implicit definitions) *constitute* the content of the term (in the same way that he denies that ‘BACHELOR → UNMARRIED MAN’ is constitutive of the content of BACHELOR).

What Fodor says is that having the concept of conjunction is “being able to think conjunctive thoughts...which is to say that it’s being able to think thoughts whose syntax is conjunctive and whose truth depends on the truth of their constituents in the familiar way” (2004a: 46).

What does this mean? Well, we know that for Fodor, having the concept TIGER doesn’t depend on being able to sort tigers from non-tigers, nor does it depend on being disposed to draw certain inferences involving TIGER (such as ‘TIGER → ANIMAL’). Rather, it depends on being able to think tiger-thoughts—that is, it depends on having a concept that means *tiger* (refers to tigers). In the same way, Fodor is now proposing that having the concept AND doesn’t depend on being disposed to accept the canonical AND-involving inferences (its introduction and

---

<sup>29</sup> He says so: “[Prinz & Clark] seem to think that I have to think that AND means what it does because it’s nomically connected to AND-ness. I don’t think that; God only knows why they think I should” (2004b: 105, fn. 7). Cf. Prinz & Clark (2004).

elimination rules). Rather, it depends on being able to think conjunctive thoughts—that is, it depends on having a concept that means *and*. (Since concepts like AND are non-referential, meaning in this case is not reference.) The next section suggests a number of problems for Fodor’s new proposal.

### **2.5.2. Problems with ‘pure’ informational atomism**

*Where do truth conditions come from?*

At the very least Fodor’s account leaves unclear the question of how the content of a logical connective such as AND is constituted. It is not enough to say that AND “expresses conjunction” or that it “means *and*”, any more than it would be sufficient to characterize the content of TIGER by saying that it expressed *tigerhood* or that it meant *tiger*. The question is *how* the concept manages to do this, that is, how its content is constituted. In the case of TIGER, informational semantics offers an answer: the content of TIGER is constituted by nomological links to tigers (or, more precisely, nomological links to the property that tigers instantiate qua tiger). In the case of AND, Fodor doesn’t offer an answer. If it’s in virtue of the truth conditions of conjunction that AND has its content, as Fodor seems to suggest, then the question becomes: Where do these truth conditions come from? Note that this is the psychological question, not the metaphysical one. The question is: In virtue of what does the item of Mentalese AND have the truth conditions that it does? Similar questions can be asked in respect of the contents of other logical terms. It is not enough to say that NOT is that concept which allows us to think negated thoughts. Fodor must also answer the question of how NOT comes to have the truth conditions that it does. If it’s not in virtue of the canonical inferences that it enters into, it’s difficult to see what the answer could be. This is a question we will return to in chapter 3.

*How are possession conditions accounted for?*

Apart from these considerations, there are independent reasons to think that pure informational atomism cannot be upheld. These reasons have to do with the question of possession conditions on concepts.

Fodor's basic position regarding possession conditions is that "satisfying the metaphysically necessary conditions for having one concept *never* requires satisfying the metaphysically necessary conditions for having any other concept" (1998a: 13–14, original emphasis).

This can't, strictly speaking, be correct. On Fodor's view, concepts are individuated by their content together with the mode of presentation. For him, the metaphysically necessary conditions for having a concept are therefore (i) being locked to the property that the concept expresses and (ii) doing so under a particular mode of presentation.

There are thus two kinds of situation where satisfying the metaphysically necessary conditions for having one concept might require satisfying those for having another concept: (i) cases of concepts with complex contents, and (ii) cases of concepts with complex modes of presentation.<sup>30</sup>

What would it mean for a concept to have a complex content? Consider Dretske's variation on Putnam's famous Twin Earth thought experiment.<sup>31</sup> Suppose that on a pre-scientific Twin Earth, it turned out that the stuff called 'water' actually consisted of two different chemicals, H<sub>2</sub>O and XYZ, with the same phenomenal properties. A sample of Twin Earth water might therefore consist of H<sub>2</sub>O, XYZ, or

---

<sup>30</sup> There is also a third possibility, that of concepts with both complex contents and complex modes of presentation. This is the situation with phrasal concepts. The content of a phrasal concept such as BLACK CAT is composed from the contents of its constituents (that is, from the content of BLACK and the content of CAT). The mode of presentation 'BLACK CAT' is structured in a corresponding way (a token of Mentalese, BLACK CAT, intended to be read as a structural description—that is, as composed out of the Mentalese tokens BLACK and CAT). This is in line with Fodor's position, of course, since atomism is only a claim about lexical concepts, not phrasal concepts. See Fodor (1990a: 58) and also Woodfield (1986: 337).

<sup>31</sup> Dretske (1981: 225f). See also Fodor (1990b: 104, 1994: 32) for discussion. The original Twin Earth thought experiment appears in Putnam (1975).



some mixture of the two. A person would acquire the concept WATER by getting locked to H<sub>2</sub>O or XYZ (or both). This is different to cases of mistaken identification (for example, if brown cows seen from a distance reliably cause us to token HORSE). Fodor deals with mistaken identification cases by noting that they are dependent (asymmetrically) on cases of correct identification. By this he means that the relation between distant brown cows, say, and tokenings of HORSE is asymmetrically dependent on the relation between horses and tokenings of HORSE—that is, if horses didn't cause tokenings of HORSE, then distant brown cows wouldn't either, but not vice-versa. This explains why HORSE expresses the property *horse*, and not the property *horse-or-distant-brown-cow*.<sup>32</sup> Notice that, in the present case, the person's nomological links to H<sub>2</sub>O and XYZ are not asymmetric: the nomological link to one substance is not asymmetrically dependent on the link established with the other substance. In such a situation, what would be the content of the individual's WATER concept? It would seem to be a complex content, equivalent to the content of the concept H<sub>2</sub>O-OR-XYZ. The concept JADE might in fact work like this, since (as philosophers know well by now) the substance it refers to can be either of two chemically distinct minerals, jadeite or nephrite (or a mixture of the two). This would strictly speaking be a break-down of atomism, albeit a relatively innocuous one.<sup>33</sup>

Consider now the case of concepts with complex modes of presentation. Take as examples the (regular, terrestrial) concepts H<sub>2</sub>O and WATER. Since they plausibly express the same property, it is clear that for Fodor these concepts must be synonymous (that is, they must have identical contents—see Fodor 1994: Lecture 3; 1998a: 15). However, they still count as *different concepts* since they have different

---

<sup>32</sup> For more detailed discussion, see §3.2 below.

<sup>33</sup> Cf. Fodor (1994: 30): “though such cases occur, *it is reasonable to treat them as accidents and to regard the missed generalizations as spurious*” (original emphasis). Note also that, as Fodor points out, in such cases WATER (or JADE) ceases to be a concept of a natural kind. See also McGinn (1989: 192). One could also see concepts with complex contents as being distinct concepts with different contents but identical modes of presentation (see Fodor 1998a: 20, fn. 16).

modes of presentation. In the former case, but not the latter case, the mode of presentation is complex (it contains as syntactic constituents 'H', '2' and 'O'). It follows that one could not possess the concept H<sub>2</sub>O without having the concept HYDROGEN (or TWO or OXYGEN). This implies that satisfying the metaphysically necessary conditions for having H<sub>2</sub>O depends on satisfying the metaphysically necessary conditions for having these other concepts. Again, however, this departure from strict atomism seems relatively innocuous.<sup>34</sup> In particular, it doesn't require Fodor to accept content-constitutive conceptual relations. For while Fodor accepts that the possession condition for the concept H<sub>2</sub>O includes having the concept HYDROGEN, he can (and does) claim that this is not a part of the *content* of the concept H<sub>2</sub>O.<sup>35</sup> Idioms may be another example of concepts with complex modes of presentation. One might argue, for example, that "KICK THE BUCKET" is a complex mode of presentation carrying the same content as the concept DIE.

Another case discussed by Fodor (1994) is that of necessarily coreferential concepts, such as TRIANGLE and (closed) TRILATERAL.<sup>36</sup> Here, Fodor would seem to have two options available. Either he could consider that these concepts express the same property (in the same way as H<sub>2</sub>O and WATER are presumed to do), or he could consider that they represent different properties. In fact, he goes with the first option, that of property identity.<sup>37</sup> But if the concepts TRIANGLE and TRILATERAL express the same property, the question then arises as to what makes these distinct concepts.

---

<sup>34</sup> Fodor acknowledges this kind of case. Just after stating his strict position, quoted at the beginning of this section, he allows that there may be a few exceptions. He seems to have in mind cases like H<sub>2</sub>O.

<sup>35</sup> Compare this with Peacocke (1992), for example, who considers that possession conditions are content-constitutive.

<sup>36</sup> See also Laurence & Margolis (1999a) for a discussion.

<sup>37</sup> He states that "...if symbols that are coinstantiated in point of conceptual or metaphysical necessity are also necessarily coextensive ('triangular' v. 'trilateral'; 'water' v. 'H<sub>2</sub>O'; 'rabbit' v. 'instantiation of rabbithood'), externalist semantics bites the bullet, assumes that they are synonymous and distinguishes them by their syntax" (1994: 61). A different kind of solution is needed for 'rabbit' v. 'undetached proper rabbit part', which raises its own problems for 'pure' informational atomism (discussed later in this section). For more detailed discussion of the ontological issues, see §4.1.2 below.

Perhaps the most obvious way to account for this is with meaning postulates. This would capture the possession conditions—that is, the fact that one could presumably not possess the concept TRIANGLE without possessing the concept ANGLE, or the concept TRILATERAL without possessing the concept SIDE.<sup>38</sup> And since meaning postulates are taken to be content-constitutive, it would account for the fact that these are distinct concepts. But if Fodor wants to maintain his strict informational atomism and eschew meaning postulates, the only possibility open to him would be to propose distinct modes of presentation for the same content. In this way, although they are synonymous, they can still be different concepts. Unlike the WATER/H<sub>2</sub>O case, there is no plausible asymmetry in this case: presumably, the modes of presentation in both cases would have to be complex. This is perhaps not particularly implausible, although again it would represent a departure from strict atomism. It also raises the question of whether it is reasonable to propose complex modes of presentation (to say nothing of property identity) in all such cases.<sup>39</sup>

Suppose instead that Fodor had chosen the other option, that *triangle* and *trilateral* are distinct properties, and therefore that the contents of TRIANGLE and TRILATERAL are different. This choice is perhaps more reasonable from an ontological point of view. Thus, Sober (1982) has argued persuasively that certain necessarily coextensive properties can confer *different* causal powers on their instances, and hence should be regarded as distinct properties. Notice, however, that if we accept that these two concepts have different possession conditions, Fodor would still have to posit complex modes of presentation in this case, a departure from strict atomism. He also has to face another difficulty, by the way, which is how

---

<sup>38</sup> We are assuming here, of course, that TRIANGLE and TRILATERAL are *primitive* concepts. If they are complex concepts, then no problems arise, but in this case the discussion in the text is still relevant, since it is still legitimate to imagine what the situation would be for the corresponding primitive versions of these concepts.

<sup>39</sup> I am thinking along the following lines here. Take distinct coreferential proper names, say, 'Cicero' and 'Tully' express the same property (*being Cicero/Tully*), but CICERO and TULLY can be distinct concepts. Do we really want to be forced to say that one or both of these concepts has a complex mode of presentation?

asymmetric dependence can be maintained. For, as Laurence & Margolis (1999a) point out, if there is a nomological link between TRIANGLE and *triangle*, there must also be a nomological link between TRIANGLE and *trilateral*. But neither seems to be asymmetrically dependent on the other, since it is necessary that both laws hold. It therefore seems impossible to distinguish being locked to *triangle* from being locked to *trilateral*. There may be a response to this problem. For although it appears that there are two laws here: a TRIANGLE–*triangle* law and a TRIANGLE–*trilateral* law, this may not in fact be the case; one may be a universal truth rather than a law.<sup>40</sup> To see why, we can adapt an argument from Sober (1982). Suppose that we are equipped with a (perceptual) angle detector, which sustains the TRIANGLE–*triangle* link. And suppose that we are also equipped with a (perceptual) edge detector, which sustains the TRILATERAL–*trilateral* link. Then although there is also a necessary TRIANGLE–*trilateral* link and a necessary TRILATERAL–*triangle* link, these are universal truths rather than laws. Notice that asymmetric dependence still holds. Suppose, for example, that we knocked out our edge detector. Then triangles/trilaterals would still cause us to token TRIANGLE, but would no longer cause us to token TRILATERAL. It is relatively clear in such a situation that it is *angles* that we are detecting, and therefore that the pertinent nomological law is between TRIANGLE and *triangle*, rather than between TRIANGLE and *trilateral*, even though it is true that trilaterals always cause tokenings of TRIANGLE.

Cases of concepts whose application appears to be context-restricted also present challenges in accounting for possession conditions. Take, for example, the concept ADDLED (discussed in Fodor 1998a: 54f; see also Quine 1973/1976). Suppose we follow Fodor in assuming that the concepts ADDLED and SPOILED both mean *spoiled*, and that the difference between the two is that the former is context-restricted to eggs. That is, the possession condition of the former, but not the latter, includes having the concept EGG. How would a possession condition such as this arise? The

---

<sup>40</sup> For further discussion of the distinction between universal truths and laws, see Dretske (1977).

most obvious solution would be to say that there was a meaning postulate attached to the concept—something like ‘ADDLED( $x$ )  $\rightarrow$  EGG( $x$ )’. Meaning postulates attached to concepts are normally taken to be content-constitutive (after all, it’s precisely *meaning* that they are supposed to be about), but as we have seen, Fodor’s strict atomism requires that he reject content-constitutive inferences such as meaning postulates. What other alternatives are there? It could be that the mode of presentation itself is complex and contains EGG (analogous to the case of H<sub>2</sub>O considered above). But there is no reason to think this would be the case, and in fact it would be a rather contrived way to ensure that the possession condition held. What Fodor seems to have in mind is something more akin to a selectional restriction. However, introducing selectional restrictions on Mentalese items would involve adding complex machinery to the theory, which is perhaps best avoided unless there are compelling reasons to do so. And it is worth bearing in mind that, at some point, selectional restrictions have to be given some explanation. It is natural to assume that selectional restrictions on natural language items are to be explained with reference to their meaning (that is, with reference to the content of the concept they express). If we instead propose that the same selectional restrictions apply to the corresponding concepts, it is difficult to see what can ultimately explain these facts.

Why does Fodor postulate that the Mentalese items SPOILED and ADDLED have the same content, rather than saying that they just mean *different things* (viz., *spoiled* and *addled*)? Then there would be no need to invoke contextual restrictions. The reason has to do with cross-linguistic considerations. Fodor is concerned about cases where one language has a single unambiguous word, the translation of which in another language depends on the context (see Fodor 1998a: 54f). Such a situation, which is rather common, might be taken to imply that the word in the first language is polysemous, and this would undermine Fodor’s claim that lexical concepts are atoms.

But Fodor’s move in any case cannot deal with the full range of cases. Consider the following case, also rather common cross-linguistically. Burmese has a verb

'hja' which is used to mean either *borrow* or *lend*.<sup>41</sup> How will Fodor deal with this case? If he adopts the same approach as with 'spoiled' and 'addled', he will have to claim that the English words 'borrow' and 'lend' are both synonymous with 'hja', but are context-restricted. But surely Fodor wouldn't want to deny the existence of concepts BORROW and LEND, expressing the properties of *borrowing* and *lending*. Such an approach also raises the question of what the contexts in question would be, and how the relevant possession conditions could be formulated. Alternatively, if Fodor claims that the three words just mean different things (that is, they express the concepts HDA, BORROW and LEND), then he cannot capture the semantic relations between them, or the fact that accurate translation is possible.

A further difficulty is how we can distinguish between being locked to the properties *hja*, *borrow*, and *lend*. For it is necessarily the case that any event instantiating one of these properties also instantiates the other two. Asymmetric dependence cannot therefore distinguish between them.<sup>42</sup>

As in the triangular/trilateral case considered above, meaning postulates can be employed in all these cases to maintain a broadly atomistic approach while also capturing the relevant semantic relations. In the case of ADDLED, we can if we wish posit a (content-constitutive) meaning postulate 'ADDLED(*x*) → EGG(*x*)', which would account for the possession condition, while maintaining the atomistic position that 'spoiled' and 'addled' express different (non-synonymous) concepts. In the second case we can again maintain the atomistic view that 'hja', 'borrow' and 'lend' express different (non-synonymous) concepts. We can also capture the relevant semantic relations. A solution might be along the following lines. It would seem that 'hja' expresses a concept, HDA, which is underspecified with regard to directionality/deixis. Such an analysis is supported by the fact that there is a whole class of similar words in Burmese, including 'θin' (study/teach) and 'ju' (bring/take).

---

<sup>41</sup> This is rather like 'rent' or 'let' in English, which can mean 'rent [/let] to' or 'rent [/let] from'.

<sup>42</sup> Laurence & Margolis (1999a: 69 fn. 86; 2003b) make a similar point concerning how to distinguish pairs of concepts such as BUY and SELL.

In usage, the linguistic or pragmatic context can give further relevant specification. The semantic relations between HDA and BORROW/LEND and other similar items could be captured via meaning postulates such as 'HDA → BORROW OR LEND'.<sup>43</sup> A similar situation arises with the concept SIBLING, not lexicalised in some languages (for example, French), or with the concept AUNT-OR-UNCLE, not lexicalised in English (see Sperber & Wilson 1998: 185). In either case we have a more general concept with two sub-concepts specified with regard to gender, and the semantic relations in each case could be captured with a meaning postulate attached to the more general concept. Since such meaning postulates are content-constitutive, this mechanism also allows us to readily distinguish between concepts expressing co-instantiated properties in a way that asymmetric dependence cannot.

#### *Quine's inscrutability problem*

There is another problem which is related to that of necessarily coreferential concepts (such as TRIANGLE and TRILATERAL discussed above), but which raises its own difficulties for pure informational atomism. Consider the case of concepts which express properties that are necessarily coinstantiated but clearly distinct, such as Quine's famous (1960) "inscrutability of reference" example concerning RABBIT and UNDETACHED PROPER RABBIT PART. These two concepts cannot be synonymous (that is, cannot have the same content) since they are not even coextensive. The problem for informational semantics is that in spite of this they are necessarily coinstantiated (that is, any instance of a rabbit is also an instance of some undetached proper rabbit part—say, an undetached rabbit ear—whereas clearly the set of rabbits and the set of undetached proper rabbit parts are distinct). Fodor (1994: Lecture 3) discusses this problem,<sup>44</sup> and concludes that no purely externalist semantics can deal with it. His solution is to bring in elements of inferential role semantics. He proposes

---

<sup>43</sup> I am not proposing this as a serious linguistic analysis. As always, the decision to propose a meaning postulate attached to a concept is to be made on empirical grounds, and would need more detailed justification than I am offering here.

<sup>44</sup> See also Gates (1996), Brook & Stainton (1997), Ray (1997) and Wakefield (2003) for discussion of the problem and Fodor's proposed solution.

selecting a ‘test predicate’ which can be used to determine which concept (RABBIT or UNDETACHED PROPER RABBIT PART) individuals are in fact tokening. Consider a test predicate such as EAR. Fodor’s proposal is that individuals who are tokening UNDETACHED PROPER RABBIT PART (but not those who are tokening RABBIT) will in some cases assent to ‘*x* is a rabbit and an ear’, since something can be both an ear and an undetached proper part of a rabbit, but nothing can be both an ear and a rabbit. This proposal assumes that Mentalese has a symbol for predicate conjunction, which is not unreasonable. The difficulty is finding a way to resolve the referential indeterminacy concerning the natural-language word for predicate conjunction itself. If we can’t do this, we can’t be sure that the individual whose dispositions we’re tapping means AND by ‘and’ (they might mean OR, for example, or some non-standard connective). Fodor proposes that the subject’s inferential dispositions unambiguously determine the content of the concept they are expressing by ‘and’. That is, the content of logical terms such as predicate conjunction is constituted by their inferential role.<sup>45</sup>

It is important to note that this, for Fodor, is a metaphysical problem. So in proposing a solution to this problem which relies on the dispositions of individuals, he is making a metaphysical claim—viz., that there is a fact of the matter as to whether RABBIT means *rabbit* and not *undetached proper rabbit part*, and that such dispositions are relevant not at the epistemological level, *but because they are (partly) constitutive of the content of RABBIT*. This is another case, then, where Fodor’s general principle—that satisfying the metaphysically necessary conditions for having one concept *never* requires satisfying the metaphysically necessary conditions for having any other concept—breaks down. In this case, a possession condition for the concept RABBIT includes (in addition to having some or other concept used as the ‘test’) having a concept of predicate conjunction. This solution is

---

<sup>45</sup> Strictly speaking, Fodor could claim that inferential dispositions can be used to unambiguously identify the content of a logical term, but deny that such dispositions are constitutive of the content of the term. See chapter 3 for relevant discussion.



not available, of course, if one adopts ‘pure’ informational atomism as Fodor now proposes.<sup>46</sup>

So for all these reasons, it seems that we *still* can’t do without meaning postulates. In which case, perhaps we should just accept the proposal that certain logical properties of concepts are captured by (content-constitutive) meaning postulates.

## 2.6. An alternative way forward?

### 2.6.1. *The problem*

All this leaves us in a difficult position. First, we have argued that we cannot justify making a principled distinction between the logical and non-logical vocabularies. However, content-constitutive inferences (meaning postulates) still seem to be the best way to account for the content of the logical vocabulary, and by extension for the logical properties of some other words that are not purely logical. I have proposed that we should make a distinction between logical and non-logical *content*, rather than between logical and non-logical *vocabularies*. Logical content is to be accounted for via meaning postulates, while non-logical content can be accounted for via informational atomism.<sup>47</sup> But such a proposal would seem to immediately run into the difficulties discussed above concerning inferential role semantics: we must find a principled way of determining which of the inferences an expression enters into are content constitutive, and which are not, and we must do so in a non-circular

---

<sup>46</sup> We will see later, however, that a solution to this problem falls naturally out of the framework that I will propose. Basically, we have inbuilt mechanisms for acquiring concepts for animal kinds such as RABBIT, on the basis of an animal-kind concept template which incorporates certain (content-constitutive) meaning postulates, such as ‘ $\varphi$  RABBIT  $\psi \rightarrow \varphi$  ANIMAL  $\psi$ ’ (see §4.3.2 below). By contrast, we have no inbuilt mechanisms for acquiring concepts such as UNDETACHED PROPER RABBIT PART, and although we could acquire such concepts reflectively (through reading Quine, say), the fact that they do not have identical meaning postulates attached (undetached proper rabbit parts are not animals, nor would we take them to be) will ensure that the content of these concepts is distinct.

<sup>47</sup> I am not claiming that meaning postulates and informational atomism are exhaustive of the mechanisms that constitute conceptual content. For example, accounting for so-called ‘procedural meaning’ (see Blakemore 2000, Carston 2002: §2.3.7) presumably requires some other mechanism.

way. Or to put it somewhat differently, we must be able to provide a principled un-question-begging distinction between those aspects of content that are to be considered logical, and those that are to be considered non-logical. In the following sections, I propose a solution.

### 2.6.2. *Apriority and analyticity*

There is a venerable philosophical tradition which gives an account of a priori knowledge in terms of content-constitutive inference (or implicit definition). According to this account, since certain inferences entered into by a concept are constitutive of its content, then these inferences must be valid (that is, they are analytic).<sup>48</sup> Moreover, since these inferences are analytic, we must be justified in asserting their validity, and this explains their apriority. Or alternatively: since certain statements (meaning postulates) define the meaning of certain of their constituent terms, then these postulates must be true, hence analytic. And since they are analytic, we must be justified in holding them true.<sup>49</sup>

Those who want to deny the implicit definition account of a priori knowledge therefore have two strategies available to them. They can either deny the relation between implicit definition (/content-constitutive inference) and analyticity, or they can deny the analytic explanation of a priori knowledge.

Horwich (1992) adopts this latter strategy. He considers the possibility that the language faculty may contain certain meaning postulates, such as the following:

(23)  $\text{bachelor}(x) \rightarrow \text{unmarried}(x) \wedge \text{man}(x)$

(24)  $x \text{ caused } y \rightarrow x \text{ preceded } y$

According to Horwich, these meaning postulates are “transmitted to that area of the brain in which beliefs are stored” (1992: 100). He suggests that it is then a simple matter to characterize a notion of analyticity (what he terms ‘I-analyticity’) which is

---

<sup>48</sup> Block (1993) terms this the ‘Plausible-Sounding Principle’ (and argues against it).

<sup>49</sup> See Horwich (1998: Chapter 6) for discussion of these issues; see also Peacocke (2004).

completely determinate in its application, and so immune from Quine's arguments.

He suggests the following characterization:

...a sentence is analytic in a person's I-language at time *t* if and only if it is a consequence of the principles of that person's language faculty at time *t* that this sentence be taken as true regardless of evidence... (Horwich 1992: 101)

By invoking a determinate notion of analyticity, Horwich is able to keep meaning postulates without having to associate the content of a concept with its entire inferential role. Content is associated with those inferences which are I-analytic, and this avoids the problems of holism which Fodor raised. The connection between I-analyticity and a priori knowledge is severed, because Horwich proposes that meaning postulates could arise or be revised in light of experience, in which case clearly no claim for the apriority of meaning postulates can be made.

There are some problems, however, with Horwich's account. First, given the kind of naturalistic, psychological project we are here engaged in, the language faculty doesn't seem to be the obvious place to locate these facts. From this perspective, it is certainly not a fact about the *English words* 'cause' and 'effect' that causes precede their effects. Rather, it is a nomological necessity, a fact about *causes* and *effects*. The question is then *where* and *how* this fact is represented in the brain. It seems we must therefore say that this is not a linguistic fact but a conceptual fact. In the same way, it is not a fact about the English word 'kill' (or about corresponding words in other languages) that killing someone entails that they die. It is a fact about *killing*, captured by properties of our concepts KILL and DIE.

In contrast with Horwich (1992), Boghossian (1993, 1994, 1997) adopts the first strategy, denying the relation between implicit definition and analyticity. He points out that for an inference to be content constitutive, it need not be analytic.<sup>50</sup> An

---

<sup>50</sup> Horwich (1998: 143) also discusses this point. In fact, Horwich (1998) seems to deny *both* the analytic explanation of a priori knowledge and the connection between content-constitutive inference and analyticity.

inference may be content constitutive for a subject provided the inference is *regarded as valid* by that subject. But regarding an inference as valid in no way requires that the inference actually *is* valid. *A fortiori*, it does not require that the inference is valid in virtue of the meanings of its constituent terms. We can imagine, for example, a pre-scientific community for whom the concept WHALE has its content constituted (in part) by the inference ' $\phi$  WHALE  $\psi \rightarrow \phi$  FISH  $\psi$ ', which members of this community take to be valid. This inference would be content constitutive for members of this community, even though the inference is not in fact valid, and therefore not analytic.

It is at least possible, then, to dissociate the notion 'content-constitutive inference' from the notion 'analytic inference'. But it is still necessary to say on what basis an inference is to be regarded as content constitutive. What is required is some means of cashing out 'content-constitutive inference' which avoids Quinean problems. Quine himself, in the paper "Carnap and logical truth" (1954/1976), regarded obviousness as one possible basis in the case of logical principles, as Boghossian (1994: 120) notes. However, it is clear that content-constitutive inferences do not necessarily reduce to 'obvious' inferences in the general case.<sup>51</sup> This is clear from the fact that there are all kinds of obvious inferences that are presumably not content constitutive. For example, the inference from 'x is a dog' to 'x is not a penguin' is obvious, but we would not propose that ' $\text{DOG}(x) \rightarrow \neg\text{PENGUIN}(x)$ ' should be added to the meaning postulates attached to DOG. So we need some other means for cashing out 'content-constitutive inference'. Here is a possibility.

## 2.7. Psychosemantic analyticity

Suppose that we follow Fodor (1975) in regarding content-constitutive inferences as mentally-represented meaning postulates, understood as inference rules. We will

---

<sup>51</sup> See also Horwich (1992: 104) on this point.

then have a uniform account of the logical properties of the so-called “logical” and “non-logical” words. The only difference between the two would be that the content of a logical concept such as AND is exhausted by its associated inference rules, whereas the inference rules associated with a referential concept such as DARK do not exhaust its content.<sup>52</sup> In both cases the inference rules constitute, at least in part, the content of the concept.

Adopting this approach leads us, like Boghossian, to deny the relation between content-constitutive inference and analyticity. The reason is that content-constitutive inferences are now to be conceived of psychologically—and while the majority of our psychologically represented inference rules are no doubt veridical, this is by no means necessary. There can be plenty of cases, such as the ‘ $\varphi$  WHALE  $\psi \rightarrow \varphi$  FISH  $\psi$ ’ example we imagined earlier, where our meaning postulates are not in fact veridical, and hence fail to be analytic.

What the approach we are proposing amounts to is this. We are specifying some property other than analyticity that can be used to determine *which* inferences are content constitutive, thereby avoiding Quine’s objections. Let us call this property ‘psychosemantic analyticity’. Then an inference will be constitutive of the content of a concept just in case that inference is psychosemantically analytic—that is, just in case the inference is associated (as a meaning postulate) with the concept in question. Of course, this notion of psychosemantic analyticity will not do the work that philosophers wanted from the traditional notion of analyticity. In particular, it is of no use in giving an account of a priori knowledge—since there is no guarantee that meaning postulates are veridical, it follows that meaning postulates cannot serve as a justification for a priori knowledge. But the present enterprise is not intended to salvage the notion of analyticity, it is rather to develop a related notion as a way to defend content-constitutive inference from Quinean objections, and therefore develop a psychologically more plausible account of conceptual content.

---

<sup>52</sup> Note that, *pace* Horwich, there is no suggestion that these meaning postulates inhabit the language faculty.

Whether or not an inference is psychosemantically analytic is an empirical question. In order to determine whether a given inference is psychosemantically analytic, hence content constitutive, we have to ask whether this inference is governed by an inference rule attached to a particular concept—that is, whether the inference is part of our deductive device. This can be discovered using standard techniques of empirical investigation, as discussed in §2.2 above.

We must also consider whether the present account of content-constitutive inference can avoid the apparent circularity identified by Fodor in the standard inferential role account (see §2.4.4 above). The question is, even if meaning postulates can give an account of the inferential relations of AND, a view that we are endorsing, can they characterize its content in a non-circular way? Meaning postulates are by definition content-constitutive. But Fodor's point is that some non-circular way must be found to formulate these postulates, and he has argued that this is not possible. He further notes (2004a: 41) that just because it may be possible to provide introduction and elimination rules for an expression, this does not itself provide support for a conceptual role account. What he suggests would provide such support are cases where grasping a concept consists in grasping such rules. In chapter 3, I develop precisely such an account, and show that it is not vulnerable to Fodor's circularity argument.

As I have said, the question of whether a given inference is supported by a meaning postulate or is represented in encyclopaedic knowledge is an empirical question which is to be determined by psychological investigation. However, the account presented above suggests a number of constraints on the kind of inferences that can be expressed by meaning postulates. First, if we possess a particular meaning postulate, it follows that, intuitively at least,<sup>53</sup> we will take the rule to express (at least a nomological) necessity. To see why, suppose that we have the

---

<sup>53</sup> I introduce this caveat to cover other issues (such as when we are exposed to philosophical thought experiments about robot cats, or biological theories about the status of whales), to which we revert in chapter 4.

meaning postulate ' $\varphi \text{ CAT } \psi \rightarrow \varphi \text{ ANIMAL } \psi$ '. Then tokening ' $\varphi \text{ CAT } \psi$ ' will (*ceteris paribus*) cause the associated meaning postulate to be activated, thereby tokening ' $\varphi \text{ ANIMAL } \psi$ '. This means that we will not be able to think about cats without thinking about them *as* animals. We will therefore take the inference from 'cat' to 'animal' to be valid, and hence believe that 'cats are animals' is (at least nomologically) necessary. This rules out the possibility that certain inferences (for example, ' $\text{GERBIL} \rightarrow \text{PET}$ ', which we would presumably not take to be necessary) are psychosemantically analytic.

But note that the converse is not true: the fact that we take some relation to be (nomologically) necessary is not in itself evidence that it is represented as a meaning postulate. It seems likely that the majority of such cases will be represented not in the form of inference rules, but in the form of propositional encyclopaedic knowledge. And propositional knowledge is not psychosemantically analytic—it is not attached to the relevant concept in the form of an inference rule, it may not be reliably activated whenever a representation containing the relevant concept is tokened, it will not support spontaneous inference, and hence it will not be content constitutive.

I would like to follow Sperber and Wilson in suggesting another constraint, which is that meaning postulates can only take the form of elimination rules, not introduction rules. This constraint is suggested by the following considerations. One obvious advantage of elimination rules is that they can be triggered by the presence of particular concepts in working memory. Elimination rules therefore make a significant contribution to cognitive efficiency. They also contribute to efficiency by avoiding the duplication of information in long-term memory. Thus, having meaning postulates of the form ' $\varphi \text{ X } \psi \rightarrow \varphi \text{ ANIMAL } \psi$ ' for a range of animal kinds X means that not only the fact that X is an animal, but also all information derivable from this fact, need not be stored under each animal kind concept. Compare this with introduction rules, such as *or*-introduction. For these rules, there is no specific trigger for their application, which means that unless constrained in some way they

would apply indiscriminately to any pairs of assumptions in working memory. Thus, *or*-introduction, say, will validly generate a never-ending stream of conclusions of the form  $p \vee q$ . This may not be problematic in an informal system, but it creates serious problems for any formal characterization of our deductive reasoning abilities in which such introduction rules are incorporated. Even if this problem can be addressed, a second problem is that the conclusions derived by introduction rules are in any case trivial in the following sense: conjoining any arbitrary proposition to an assumption through (for example) *or*-introduction does nothing to improve an organism's understanding of the world. Based on these considerations, Sperber & Wilson (1995: Chapter 2) propose that the only inference rules attached to concepts are elimination rules—that is, there are no introduction rules (such as  $\vee$ -introduction). In the next chapter, we will look in more detail at how this approach might work.

## 2.8. Conclusion

This chapter has argued for two things. The first is that meaning postulates are indispensable for a variety of reasons, including in accounting for the logical properties of concepts, and for their possession conditions. The second is that if we understand meaning postulates as mentally-represented rules of inference, then we can avoid the apparently fatal consequences of Quine's classic argument against such a position, by providing a psychological basis for the distinction between content-constitutive inferences and the rest.

There is an unanswered question here, however. In arguing for a psychologically plausible account of mental inference, we have suggested not only that inference rules are required, but also that they should be restricted to elimination rules. As we saw, there is psychological justification for eschewing introduction rules. But how, then, are we to provide an account of how concepts get their logical content? Traditionally, inferential role semantics has considered that the canonical elimination *and introduction* rules are the basis on which logical content is constituted. Without



introduction rules, it would appear that we are lacking a complete account of logical content. In the absence of such an account, it might be tempting to cast in one's lot with Quine, and try to argue that although we may require mentally represented inference rules in order to account for deductive reasoning, such rules need not be content constitutive.

It is to these considerations that we turn in the next chapter. By looking in detail at the nature of conceptual content, in particular in the case of logical terms, we will see that elimination rules can be the sole basis on which logical content is constituted, and that meaning postulates cannot *but* be content constitutive.

### 3. Content

#### 3.1. Introduction: what is conceptual content?

This chapter will not provide an exhaustive or comprehensive account of the various theories of conceptual content. What I will do is start with a fairly general and (relatively) untendentious discussion of content, albeit one that presupposes a representational theory of mind,<sup>1</sup> and then explore whether this can help shed some light on various issues raised in chapter 2.

It is a curious fact about the world that some things (sentences, thoughts, propositions) are *about* other things (tables, chickens, even other sentences/thoughts/propositions). Thus the sentence ‘Boris the cat is black’ is about Boris the cat; and the thought ‘chickens are tasty’ is about chickens; and the proposition ‘ $\text{John}(x) \rightarrow \text{greedy}(x)$ ’ is about John. This property that some things have of being about other things is known as ‘intentionality’. We are in need of some explanation, since it is not immediately obvious *how* something can be *about* something else. It’s not, for example, likely that intentionality is a fundamental property of things in the same way as mass or charge or spin.

There is clearly a link between intentionality and representation. The sentence ‘Boris the cat is black’ conveys a certain piece of information (represents a certain state of affairs), namely that Boris the cat is black. This same information could, it seems, also be conveyed in other ways. For example, a photograph or a drawing of Boris the cat could convey the same information (it too represents a certain state of affairs). There are, however, important differences between these two modes of representation. A fundamental difference is that while the photograph or drawing represents iconically (that is, it resembles what it represents),<sup>2</sup> the sentence

---

<sup>1</sup> See §1.2.1 for a discussion.

<sup>2</sup> Actually, things are a little more complex. Although photographs are iconic representations, in the sense that they resemble what they represent, they do not necessarily represent what they resemble. That is, their content is fixed not by resemblance but by their causal properties—a photograph of Boris is not a photograph that resembles Boris, however closely, but a photograph caused by Boris (so it needs to have been Boris that was sitting in front of the camera when the photograph was taken, and

(/thought/proposition) represents in an indirect or coded way (that is, by means of abstract meaning relations). This fundamental difference gives rise to a number of specific differences between these modes of representation.

One significant difference concerns the amount of information that is conveyed. A sentence, thought or proposition conveys highly specific information (for example, the information that the cat is black). A photograph, on the other hand, cannot convey one piece of information (the colour of the cat, say) without also simultaneously conveying an indefinite number of other pieces of information (the cat's size, shape, orientation, and so on). Dretske (1981: 137) calls this a distinction between 'digital' and 'analogue' forms of representation:

...a signal (structure, event, state) carries the information that  $s$  is  $F$  in *digital* form if and only if the signal carries no additional information about  $s$ .... When a signal carries the information that  $s$  is  $F$  in analogue form, the signal always carries more specific, more determinate, information about  $s$  than that it is  $F$ .

This is important. Intentionality is the property that something has of being about something else. Sentences, thoughts and propositions have intentionality (so the thought that Boris is black is a thought *about* Boris). Photographs do not have intentionality in the same way (while a photograph of Boris might be about Boris, it might also be about many other things: cats in general, quadrupeds, fur, a table which happens to have a cat sitting on it, a room, and so on).

Another important difference concerns the possibility for conveying false information (misrepresentation). A photograph's meaning is tied to its information

---

not, say, Boris's identical twin). Compare this with drawings. Drawings are also iconic representations, but again their content is fixed not by resemblance but (postmodernist art critics notwithstanding) by reference to the intentions of the creator: a sketch intended to be of Boris is a representation of Boris, and not of Boris's twin, *even if Boris's twin posed for the artist* (to help get the shape of the face right, say). The sketch may not even resemble Boris (perhaps the artist is intentionally abstract or just not very skilled). The classic text on the different types of representation is Peirce (1931–1935), although the specific proposals that it makes have been rejected by many later theorists. See Goodman (1976).

content, which it records through a deterministic process of photons striking a light-sensitive medium, and so on. A photograph can only carry the information that  $p$  if the state of affairs  $p$  in fact pertains (/pertained). So a photograph of Boris the cat can only carry the information that Boris is black if in fact Boris actually is (/was) black. If the photograph is deliberately manipulated so that it depicts Boris as being white, it does not carry the information that Boris is white, since there is no such information for it to carry, and the photograph therefore does not mean that Boris is white. Thoughts and other intentional structures are not like this. The thought that Boris the cat is white can be false or true, depending on how the world is constructed, and the thought represents this state of affairs whether or not it happens to be true. So while representation and intentionality are related notions, they pull apart in important respects. (For detailed discussion of these issues, see Dretske 1981, 1986, Grice 1957, and Woodfield 1986.)

Let us now consider the more specific case of the propositional attitudes. As the name suggests, propositional attitudes are attitudes that we can take towards propositions. Examples of propositional attitudes are believing that a glass of water is on the desk before me, desiring to drink some water, intending to raise the glass to my mouth, and so on. In each case, there is an attitude (belief, desire, intention) towards a proposition (expressed by a 'that...' or 'to...' clause). The proposition expresses what is known as the content of the attitude. A common way of talking about propositional attitudes, following Stephen Schiffer, is in terms of a series of 'boxes' corresponding to the various attitudes. On this analogy, the event of my believing that  $p$  comes about as a result of having in my belief box a representation that means  $p$  (the same representation appearing in my desire box would count as an event of my desiring that  $p$ , and so on *mutatis mutandis*). The analogy is rather direct: the content of the attitude corresponds to the contents of the box in question.<sup>3</sup>

---

<sup>3</sup> Note that propositions are abstract entities, so it's a representation (a token of a symbol that expresses the proposition) which appears in the attitude-box, not a proposition. See Fodor (1978). Attitude boxes are of course just a convenient metaphor. The idea behind this metaphor is that attitudes are typed with respect to their functional properties (for example, desires tend to cause action

I will assume that the representations expressing the contents of the attitudes are structured in broadly the same way as the natural-language clauses used to express them (this is the language of thought hypothesis). That is, mental representations of clauses are built up from lexical concepts, combined in accordance with the syntactic principles of the language of thought (in the same way that natural-language clauses are built up from lexical items combined in accordance with the syntax of the natural language in question). It follows that the content of a clausal representation is built up from the contents of its constituent concepts, together with the syntax. (And in fact, nothing in what follows will depend on there being a language of thought; all we need is the much less contentious assumption that thought is compositional.) The content of propositional attitudes is therefore derivative; underived content is to be located at the level of lexical concepts. There is, however, a lively debate in the literature—even among proponents of a representational theory of mind—as to whether *all* content is conceptual.<sup>4</sup>

The debate over nonconceptual content is not a new one. Philosophers from Kierkegaard to Kant have held that aesthetic ideas cannot be adequately conceptualized, and therefore that whatever content is expressed by the arts is nonconceptual (for discussion, see the papers collected in Gunther 2003). The contemporary debate revolves primarily around perceptual experience. As Peacocke (1992: Chapter 3) notes, a particular perceptual experience represents the world as being a particular way, and it does so with a precision and detail apparently beyond that which could be conceptualized. For example, the range of colours that we can perceive seems far greater than the range that we can conceptualise. Similar points can be made for other elements of percepts, such as precise shapes, luminosities, and so on. Some theorists therefore conclude that the content of perceptual experience must be nonconceptual.

---

designed to bring about the conditions for their satisfaction). Importantly, this proposal is intended to be neutral on the question of how the *content* of the attitudes is to be determined.

<sup>4</sup> See, for example, Peacocke (1983: Chapter 3, 1992: Chapter 3), Fodor (1998b, 2004c) and the various papers collected in Gunther (2003).

Whether their content is conceptual or nonconceptual, though, all intentional states have the property of semantic normativity: they represent things as being a certain way, and as we have seen above, they do so independently of whether things actually are that way. Intentional states therefore have truth conditions (or, more generally, satisfaction conditions)—sets of conditions that determine whether a particular belief is true, or whether a particular desire is satisfied. Thus, for example, the belief ‘Boris the cat is black’ is true iff the world is arranged in such a way that the individual cat in question (Boris) has the particular property in question (blackness). Similarly, my desire to drink water is satisfied iff a certain state of affairs is brought about (viz., the state of affairs expressed by the proposition in question: I drink some water).

Propositional attitudes have two important properties, then. They have propositional (conceptual) content, and they have satisfaction conditions. The question arises how these two properties are related. It is first worth noting that whatever propositional content is, it is something abstract. Propositional attitudes are instantiated because we are related in a certain way to a proposition, by having a mental representation expressing that proposition in our belief box, say. A token mental representation expresses an abstract proposition, so the content of that proposition is presumably itself something abstract. Next, notice that the satisfaction conditions of a propositional attitude reduce to the truth conditions of its propositional object: a belief is true if its propositional object is true; a desire is satisfied by making its propositional object true; similarly, an intention is realised when the truth of its propositional object is brought about. Truth conditions are also abstract. They are a characterization of how the world would have to be in order for the proposition to be true.

So, propositional attitudes have content, which is something abstract, and their propositional objects have truth conditions, which are also abstract. One obvious conclusion which might be drawn from this is that propositional contents just are truth conditions. The content of a concept would then be seen as the contribution that

concept makes to the truth conditions of propositions within which it occurs. And, indeed, this is a fairly widely-held assumption. One possible objection to proposing that propositional contents and truth conditions are identical, however, is that this rules out the possibility that different propositional contents have the same truth conditions. If we wish to leave this possibility open, then, we would need to find some abstract characterization of content that was more fine-grained than truth conditions—that is, some function from propositions to truth conditions. (For discussion of these points, see Stalnaker 1998.)

Fodor (1998a) has a somewhat different solution. He holds that the contents of the propositional attitudes are truth conditions, but also accepts that two propositions can have the same content, and therefore the same truth conditions, and yet be *different* propositions. He does this by adapting the Fregean proposal that concepts are individuated by both reference and mode of presentation. According to Fodor's adaptation of this position, concepts are individuated not only by their contents, but also by the way in which this content is presented to thought, that is, its mode of presentation. For Fodor, modes of presentation are not Fregean senses, but rather language of thought expressions. And since, on present assumptions, propositions inherit their content from their constituent concepts (plus the syntax), this explains how different propositions can have the same content, and hence the same truth conditions.

The above gives some characterization of the abstract objects that might be propositional contents. A separate question is: in virtue of what does a particular representation come to have the content that it does? In other words, how is it that an abstract content becomes attached to a token mental representation? This is the question that we raised at the end of chapter 2. One answer that has been proposed (as we saw in §1.2.1 and §1.2.6 above) is informational semantics. The difficulty we had, however, was that it was not obvious that informational semantics could work as an account of the content of the logico-syntactic apparatus. Let us first, then, look at informational semantics in more detail.

### 3.2. Informational semantics

We are assuming that a token mental representation gets its content from its constituent concepts together with their mode of combination (the syntax). The problem of how a mental representation gets its content therefore reduces to the question of how primitive concepts do so.<sup>5</sup>

According to informational semantics (as set out in Dretske's seminal 1981 book *Knowledge and the Flow of Information*), primitive concepts get their content from their nomological relations to the entities that fall under them. Exactly how this works is a somewhat complicated story, but in what follows I will try to summarize what are the essentials for our present purposes.<sup>6</sup>

As we noted above, there is clearly some sort of link between intentional content and representation. Part of the story as to how a mental state can be *about* black cats pretty clearly has to do with the fact that the mental state in question *represents* black cats. And representing black cats seems, intuitively at least, to have to do with *carrying information about* black cats. So it seems, at least *prima facie*, that the content of a mental state supervenes on the information that it carries. Since 'information' (at least in quantitative terms) is a well-defined mathematical concept, this looks to be a promising strategy for naturalizing intentional content—that is, for giving an account of intentional content in non-intentional terms. This is Dretske's project.

There are two immediate problems that any attempt to reduce content to information must deal with, as we saw above. First, the content of a mental state is determinate, whereas the information carried is not. Thus, a mental state that carries the information *A* must also carry the information *A or B* and the information *A or B*

---

<sup>5</sup> Dretske, though not Fodor, leaves open the possibility that some complex concepts get their content not from their constituents, but directly (that is, that the concept BLACK CAT, say, gets its content not from the contents of BLACK and CAT, but from its nomological links to black cats). It is not clear, though, that it is necessary—or wise—to do so. See Fodor (1990a: 58) and footnote 30 (p. 61) above.

<sup>6</sup> For a more detailed (and no doubt more faithful) summary see Dretske (1983). A very clear and concise discussion of the issues is also to be found in Woodfield (1986).



or *C*, and so on; a mental state with the content *A*, however, does not also have the content *A* or *B*. Similarly, a mental state that carries the information that the temperature is 100 degrees also carries the information that the temperature is higher than 90 degrees, whereas a mental state with the content *the temperature is 100 degrees* does not thereby also have the content *the temperature is higher than 90 degrees*. Dretske's solution to this problem is to introduce a distinction between digital and analogue forms of information carrying. Every signal carries information in both digital and analogue forms, but Dretske (1981: 137) stipulates that the most specific piece of information that a signal carries is the only piece it carries in digital form, with all other information being carried in analogue form. Dretske notes that sensory and cognitive processes can be distinguished on this basis: cognition consists in the extraction of certain pertinent information (in digital form) from the plethora of information presented by our senses (in analogue form). The content of a mental state, then, is related to the information carried by that state *in digital form*.

The second problem is how to account for false contents. Mental states have their contents independently of how the world is constructed, so the contents of mental states may be true or false. But it is not clear that informational semantics can accommodate this fact. If it is not the case that *s* is *F*, then there is by definition no information that *s* is *F*. So if the content of a mental state is related to the information carried by that mental state, it would not seem possible to account for false contents.

A number of solutions to this problem have been proposed. Dretske's own solution (1981: Chapter 8) was to propose that there was a learning period during which mental states were trained to digitize a particular piece of the range of information presented in analogue form, perhaps through some kind of feedback mechanism. In this way, a particular mental state came to be associated with a particular piece of information, so that future mental state tokens of this type inherited this particular content, whether or not these subsequent mental tokens actually carried that information. This allows for the content of a token mental representation to be identified with the information carried (during the learning

period) by the corresponding mental representation type, without it necessarily being the case that all tokens of that type (or indeed any, outside the learning period) carry that information. In this way, a token mental representation can misrepresent, and can therefore have a content that is false.

Fodor (1984a) objects to Dretske's proposal, on various grounds. First, he points out that it is difficult to see how to draw a principled distinction between the learning period and the post-learning period. He also notes that Dretske's solution can only account for the misrepresentation of learned symbols, not innate symbols. Crucially, though, Fodor also points out that any situation that gives rise to a misrepresentation  $B$  after the learning period would also have caused a similar misrepresentation (that is, a false tokening of a mental representation 'A' that means  $A$ ) if it had occurred during the learning period. How do we then declare that what was being learned was that the mental representation of type 'A' meant  $A$  rather than  $(A \vee B)$ ? This is what Fodor calls the 'disjunction problem'. Notice that Dretske can't just do without the learning period and say that the content of a token mental representation is identified with the information typically carried (in digital form) by its corresponding representation type, because this leaves him immediately open to Fodor's disjunction problem again: if a representation of type 'A' typically carries the information that  $A$  and only occasionally the information that  $B$ , then this is better explained by that representation type meaning  $(A \vee B)$ , which is the information that it carries even more typically than the information that  $A$ . In general, any apparent misrepresentation can always be subsumed by some sufficiently complex disjunction. So we are left without an account of how misrepresentation is possible.

One possible way around this problem would be to say that what a mental representation of type 'A' means is linked with the information it carries (in digital form) *in normal circumstances*. The difficulty then becomes to give a naturalistic account of what constitutes 'normal circumstances'. As Fodor (1984a, 1987: Chapter

4, 1990a) points out, however, it's not at all clear that this is possible.<sup>7</sup> The typical way of accounting for 'normal circumstances' is by appeal to teleology underwritten by Darwinian natural selection. Thus, the content of a mental representation of type 'A' can be identified with whatever it *normally* carries information about, which is a matter of what it was *selected* to carry information about. First, it's not clear that natural selection always favours veridical representations—that is, that what's selected for must ipso facto be true. Perhaps, argues Fodor, there are situations in which it is of greater survival value to an organism to represent something as false rather than veridically (think of repression of unbearable truths). Worse, it is often not clear exactly what information a mental representation type was selected to carry. Consider the famous case of the fly-eating frog, discussed in Fodor (1990a). Suppose (what is *prima facie* reasonable) that the frog tokens a mental representation of type 'F' in the presence of flies, and that this causes appropriate fly-catching behaviour. Suppose also that it's not just flies, but moving black dots in general, that cause tokenings of 'F'. Tokenings of 'F' caused by moving black dots other than flies are misrepresentations only on the assumption that in normal circumstances it's flies that cause 'F'-tokenings, and that this is so because natural selection favoured a mechanism that responded selectively to flies (so that 'F' has the content *fly*). But Fodor argues that there's another equally valid way of telling this story: supposing that in the environment in which the mechanism evolved it happened that most moving black dots were in fact flies, we can say that natural selection favoured a mechanism that responded selectively to moving black dots. Thus, when the frog tokens 'F' in response to a non-fly this may result in a case of indigestion but it is not a case of misrepresentation (the content of 'F' in this case being *moving-black-dot*). How to choose between these two versions of the story? In terms of natural selection it doesn't seem to matter. In fact, Fodor suggests that as far as natural selection is concerned, we could equally well say that the content of 'F' is *fly-or-inedible-*

---

<sup>7</sup> One of the most thorough attempts to work out a theory of this kind was by Fodor himself (1984b/1990). See also Millikan (1984).

*moving-black-dot*. All that natural selection cares about is that in the frog's environment (more accurately, in the environment in which the frog's 'F'-tokening mechanism evolved) what falls under FLY-OR-INEDIBLE-MOVING-BLACK-DOT tends to be flies. Fodor concludes from this that teleology can't provide an answer to the disjunction problem.<sup>8</sup>

Having rejected both Dretske's learning period solution and the teleology solution, Fodor then sets out his own solution to the disjunction problem (see 1990b). The basis of Fodor's proposal is a fundamental asymmetry between false tokens and true tokens. Suppose that cows cause COW-tokens, and so too (say) do some horses (distant ones in the evening light, perhaps). However, COW still means *cow* and not *cow-or-horse*. The reason for this, according to Fodor, is that those COW tokens that are caused by horses depend on the fact that there are COW tokens that are caused by cows, but not the other way round. As Fodor puts it, "noncow-caused COW tokens are *asymmetrically dependent upon* cow-caused COW tokens" (1990b: 91, original emphasis). Another way to think of this is that the nomological link between cows and COWs is in a certain sense more basic than the nomological link between distant horses in the evening light and COWs. The latter hijacks the mechanism that links cows with COWs (our cow-detector, as it might be), so if cows did not exist, the link between distant horses in the evening light and COWs would be severed, whereas if horses did not exist, cows would still cause COWs. This is the fundamental asymmetry that Fodor makes use of to try and solve the disjunction problem.

---

<sup>8</sup> There's another problem for teleological theories that Fodor raises but which I won't go into here. It's that some tokens of 'F' that are not caused by Fs are nevertheless not errors (and therefore occur even in teleologically normal circumstances). For example, thinking about frogs may lead to thinking about flies (that is, to tokening FLY). This is not an erroneous tokening of FLY even though it's been caused by something that isn't a fly (viz., by a mental representation, in this case FROG). See Fodor (1990c).

### 3.3. Problematic cases for informational semantics

A number of problems have been raised for Fodor's asymmetric dependence theory as a solution to the disjunction problem. I will not discuss these here,<sup>9</sup> but instead want to focus on a different kind of problem, which is that not all classes of concept are amenable in any case to Fodor's treatment.

As we have seen, according to Fodor's informational semantics, concepts get their content from a nomological link with the property they express. This is fine for natural kind concepts like CAT and TREE, which plausibly express properties (*cathood* and *treehood*). And it may be that this treatment is extendable to nominal kind concepts such as BACHELOR which can be seen as expressing the property *bachelorhood*, and even to artefact concepts such as DOORKNOB which Fodor (1998a) argues expresses the (mind-dependent) property *doorknobhood*. But it is not *prima facie* plausible that informational semantics could account for the content of proper names or the logico-syntactic apparatus, since CHOMSKY does not plausibly express the property of *chomskyness*, just as AND does not plausibly express the property of *andness*.

#### *Proper names*

First, consider proper name concepts. Unlike natural kind concepts, say, which express properties that any number of individuals can instantiate, proper name concepts are a species of individual concept: only one individual can ever fall under them.<sup>10</sup> This is why it does not make intuitive sense to speak of a proper name concept (such as CHOMSKY) as expressing a property (*chomskyness*). There is no property, no hidden essence, possession of which will mean that an individual falls under the concept CHOMSKY, other than (trivially) the property of *being Chomsky*, a property that necessarily only Chomsky can instantiate. For familiar Kripkean reasons, even *the author of Syntactic Structures* won't do the trick, since it is

---

<sup>9</sup> See, however, Fodor (1990b), and the papers in Loewer & Rey (1991).

<sup>10</sup> It's possible, of course, for *no* individuals to fall under a proper name concept, as is the case with empty names (SANTA CLAUS, say).

impossible for Chomsky not to be Chomsky, while it is at least possible for Chomsky not to be the author of *Syntactic Structures*.<sup>11</sup>

Fodor has at different times endorsed different accounts of the content of proper name concepts, usually without much discussion.<sup>12</sup> In Fodor (1987: 84 ff.), while noting that “The course of wisdom would be to reiterate the moral—viz., that names are a hard problem for everybody—and then to shut up and leave it alone”, he briefly sketches a variant of the description theory of names. According to this version, a concept such as CHOMSKY has as its content *the person named ‘Chomsky’*.<sup>13</sup> Importantly, this account is able to deal with Frege cases. For example, the concepts CICERO and TULLY are distinct, even though they both refer to the same individual, because they express different (linguistic) properties: the property of *being the person named ‘Cicero’* and the property of *being the person named ‘Tully’*. A well-known problem with this proposal, however, is that it implies “Cicero is named ‘Cicero’” is a necessary truth, when this is clearly not the case (he could have been called anything at all). This leads Fodor to reject this account as it stands.

Fodor’s twist is to treat proper names like demonstratives, and claim that, for example, “Cicero was bald” says “he<sub>Cicero</sub> was bald” and it presupposes that he is called ‘Cicero’. This allows “Cicero is named ‘Cicero’” to come out contingent, as it should be, since “he<sub>Cicero</sub> is named ‘Cicero’” presupposes that he is called ‘Cicero’ only in this world, not in all possible worlds. It also allows us to maintain the intuition that, since *being Cicero* and *being Tully* are the same property, CICERO and TULLY have the same meaning. What makes them distinct concepts is that they differ

---

<sup>11</sup> See Kripke (1972/1980) and the extensive literature that this gave rise to. (For some recent attempts to defend descriptivism, see Stanley 1997, Sosa 2001, and Nelson 2002; Everett 2005 argues that these attempts are ultimately unsuccessful.)

<sup>12</sup> Cain (2002: 116) suggests that Fodor endorses a causal–historical account of the content of proper names. While this is probably true of Fodor (1987) it is certainly not true of Fodor (1994), as we will see below.

<sup>13</sup> Such a theory was proposed by Kneale (1962), and has more recently been defended in Geurts (1997).

in presupposition, and hence (at least on some accounts of presupposition) have different truth conditions.

But note that this is not an informational semantic account, despite some suggestive language from Fodor (1987: 85, emphasis and bracketed sentence in original):

‘Cicero is Tully’ is informative because, although it doesn’t *say* that the guy who was called ‘Cicero’ was called ‘Tully’, it “carries the information” that he was. (For more on this notion of carrying information, see Dretske 1981 and Barwise & Perry 1981.)

What Fodor seems to be getting at is this. The sentence ‘Cicero is Tully’ expresses the proposition CICERO IS TULLY, which has the (referential) content *he<sub>Cicero</sub> is he<sub>Tully</sub>* (uninformative) but which presupposes *the person named ‘Cicero’ is the person named ‘Tully’* (informative). This is certainly not an informational semantic treatment. There is no nomological link or lawful correlation here between the concept CICERO and the information that Cicero is called ‘Cicero’ (as we have seen above, it can’t be the case that CICERO expresses the property of *being the person named ‘Cicero’*; if CICERO expresses any property, it’s the property of *being Cicero*). Rather, it seems that there could be a causal–historical link between the concept CICERO and the fact that Cicero is called ‘Cicero’. But to repeat, this is not a nomological link, and therefore this is not an informational semantic account in the sense proposed by Dretske and adopted by Fodor.<sup>14</sup> That it is a causal–historical account that Fodor has in mind is also made very clear by the fact that he explicitly likens his treatment of proper names to the treatment of demonstratives. And demonstratives in his view demand a causal–historical treatment if anything does: the only remotely plausible account of the content of the concept THAT BOOK (say) is whichever book *actually gave rise to that particular Mentalese token*.

---

<sup>14</sup> Fodor himself has been careful to stress that the causal–historical and the nomological should not be conflated, particularly in swamps. See Fodor (1994: Appendix B).

In Fodor (1994: Appendix A) we get a rather different treatment of proper names, which Fodor explicitly contrasts with his earlier account outlined above,<sup>15</sup> and which is basically an informational semantic account. Somewhere between 1987 and 1994 Fodor has realised that his earlier account outlined above account doesn't work: apart from anything else, as Fodor (1994: 112 f.) points out, it's just not the case that there is anything specially metalinguistic about names. It's true that "he is Cicero" invites the inference that he is called Cicero, but "that is a rose" can similarly invite the inference that that is called a rose, without anyone supposing that "rose" has metalinguistic properties.

According to Fodor's new account, it is plausible to assume that CICERO and TULLY carry the same information, since they express the property *being Cicero* and the property *being Tully*, which are plausibly the same property. Assuming an informational semantic account, CICERO and TULLY therefore have identical content. What, then, makes them different concepts? The normal way to proceed, as we have seen earlier, would be to propose that they have syntactically-distinct modes of presentation—that is, to say that one or other (or both) of them have complex modes of presentation. But this is implausible—why wouldn't CICERO and TULLY be syntactically primitive, just as the corresponding natural language words are? Fodor instead proposes that there must be some other formal (possibly neurological) difference between CICERO-tokens and TULLY-tokens which allows them to be type-distinct while having the same content and syntax. We will discuss what kind of difference this could be in chapter 4. All that is left to explain is *how* the concepts CICERO and TULLY come to express the property of *being Cicero/being Tully*. That is, what mechanism is it that sustains this link? Here, Fodor (1994: 118 f.) again adopts a causal-historical account, this time not as a metaphysical account of the content of proper names, but as an explanation for why the nomological link between the concept and the property holds. That is, the causal-historical properties of proper

---

<sup>15</sup> See Fodor (1994: 111): "...I do want to stress the difference between this view and (what I'll call) the Metalinguistic View, viz., that 'Cicero', but not 'Tully', means something like *is called 'Cicero'*."



names are the mechanism that sustains the nomological link between proper name concepts and the corresponding properties. But, crucially, for Fodor it is the existence of such a link, not the details of the mechanism that sustains it, which underpins the metaphysics of proper name content.

### *Logical terms*

The other class of concepts that presents difficulties for informational semantics is the logical terms (or, more broadly, the logico-syntactic apparatus). As was the case with proper names, it doesn't seem *prima facie* plausible that logical concepts get their content from a nomological link with the property they express. With proper names, this intuition stemmed from the fact that the entity they refer to seems not to be picked out by reference to any particular property that it possesses. In the case of logical concepts, it seems that they do not refer at all. It does not seem plausible, therefore, to adopt the usual informational semantic analysis and propose that AND (say) gets its content from a nomological link with a property of *andness* or *conjunction*.

The usual approach taken by informational semanticists has been to cede the logical terms to inferential role semantics. That is, it has generally been accepted that informational semantics is not the right approach for dealing with the logical vocabulary. The informational semantics dictum "meaning is reference" is fine for those concepts that refer, but for those concepts that do not refer, meaning must be constituted by something else, and inferential relations are an obvious candidate.

Recall that the metaphysical question, "what content does a particular concept have?" is answered by stating what contribution the concept makes to the truth conditions of propositions in which it occurs. In the case of logical terms, then, their content is the logical contribution they make to the propositions they occur in—that is, their logical properties. Take, for example, a subset of the logical terms, the logical connectives. The contribution that a logical connective makes to the truth conditions of a proposition in which it occurs is its particular (Boolean) function—

that is, its truth table. This is to give a characterization of the abstract objects that might be the contents of the logical connectives. The question then is to give an account of how a token mental representation comes to have this abstract object as its content. This is the question which, in the case of logical terms, semanticists have traditionally answered with an inferential role account, typically in terms of implicit definitions or possession conditions which provide introduction and elimination rules for the concept (see §2.4.4 above).

For example, Peacocke (1992: 6) proposes that the possession conditions on the concept of conjunction can be identified with the transitions of the forms in (1), which a possessor of conjunction must find “primitively compelling”.

- (1) a.  $p, q / pCq$   
b.  $pCq / p$   
c.  $pCq / q$

These, of course, are the standard introduction and elimination rules for conjunction. As such, it is easy to show that together they uniquely specify the logical properties of conjunction (see below). Peacocke’s proposal is not this, which would be trivial, but rather that *grasping the rules in (1) in the right way just is to possess the concept of conjunction*. This is to claim that there is nothing more to having the concept of conjunction than finding its canonical introduction and elimination rules (primitively) compelling. In particular, there is no need to postulate mind–world links, as there is on an externalist account of concept possession.

As we have seen in chapter 2, Fodor (2004a, 2004b) now rejects this view of the content of logical terms, which he believes to be viciously circular. Instead he proposes that having the concept AND (say) doesn’t depend on being disposed to accept the canonical AND-involving inferences (its introduction and elimination rules), but rather on being able to think conjunctive thoughts—that is, it depends on having a concept that means *and*. The details, however, are a little thin on the ground. In the following section, I develop an argument which demonstrates, in

support of Fodor's position, that possessing the concept of a logical connective need not require accepting the full set of introduction and elimination rules for that concept. This casts doubt on the inferential role view of concept possession according to which grasping the introduction and elimination rules is necessary and sufficient for possession of the concept. This still leaves open the crucial question of how a token mental representation comes to have the logical content that it does, to which Fodor did not give any detailed answer. In what follows, I will set out a psychologically plausible and independently motivated account which can provide an answer to this question. Relying as it does partly on meaning postulates, I am not sure that this account is one that Fodor would approve of, however. This chapter will close by drawing out more clearly some of the points of agreement and disagreement between my approach and Fodor's.

### **3.4. Logical connectives and their canonical inferences**

Consider the truth-functional connective ' $\wedge$ '. This connective is governed by standard introduction and elimination rules as set out in (1). Is it necessary in order for a mind to have the concept AND that it grasps these three rules? According to inferential role semantics, the answer is 'yes'. On such an account, for a mind to have AND *just is* for that mind to grasp these three rules. I will argue, however, that the correct answer to this question is 'no'. It is possible for a mind that does not grasp all of these rules to nevertheless be able to think conjunctive thoughts and therefore to have the concept AND. In particular, it is possible for a mind which does not grasp the introduction rule for conjunction to nevertheless have the concept AND. This is what I aim to demonstrate below.

First, consider why it is that grasp of the introduction and elimination rules in (1) is sufficient for grasp of AND. Given the propositions  $p$ ,  $q$ , and  $p * q$ , grasp of the rules in (2) will be sufficient for grasping ' $*$ ' as AND.

- (2) a.  $p, q / p * q$   
 b.  $p * q / p$   
 c.  $p * q / q$

Syntactically, ‘\*’ operates as a binary connective. Grasping it as AND requires ruling out other inconsistent interpretations. In standard Boolean logic, there are 15 other possibilities which need to be ruled out, as shown in (3).

(3)

		And															
<i>p</i>	<i>q</i>	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
0	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
1	0	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

The three rules in (2) are sufficient for grasp of AND as they rule out the other 15 possibilities in (3). Since the content of a logical connective is just its truth table, for a logical concept to pick out a (unique) truth table is all that is required to constitute the content of that concept. The rules in (2) pick out the truth table for conjunction as follows. Rule a. states that if two propositions are true, connecting them with ‘\*’ results in a proposition that is true. This rules out those functions that give a value of ‘0’ when both *p* and *q* have a value of ‘1’—that is, it rules out all the even-numbered functions in (3) (including function 0). Rule b. states that if a proposition of the form ‘*p* \* *q*’ is true, then the constituent proposition *p* is true. This rules out those functions that give a value of ‘1’ when *p* has a value of ‘0’—that is, it rules out functions 4–15 in (3). Finally, rule c. states that if a proposition of the form ‘*p* \* *q*’ is true, then the constituent proposition *q* is true. This rules out those functions that give a value of ‘1’ when *q* has a value of ‘0’—that is, it rules out functions 2–3 and 6–15 in (3). The only function which has not been ruled out is function 1, which is the function for ‘and’. It follows that grasping the rules in (2) is sufficient for

grasping ‘\*’ as AND. No other Boolean connective is compatible with these three rules.<sup>16</sup>

However, this in itself does not imply that grasping the three rules in (2) is *necessary* for grasping AND, even if there is no smaller set of rules which uniquely identifies ‘and’ among the possible Boolean connectives. To see why, suppose that there are other more general considerations which rule out some possible Boolean connectives. Suppose, for example, that our minds incorporate a general constraint which rules out truth-functional contradictions (that is, functions whose values are always false, regardless of the values of their inputs) as possible connectives.<sup>17</sup> Such a general constraint, together with just the elimination rules in (2) (that is, just rules b. and c.), is sufficient to rule out the 15 other possibilities in (3). Notice that the only function eliminated by rule a. which is not also eliminated by rules b. or c. is function 0 (contradiction). Since this is ruled out by our hypothetical constraint, it follows that the elimination rules alone are sufficient to uniquely identify ‘\*’ as AND. Thus, for any mind that rules out truth-functional contradictions on more general grounds, grasp of the elimination rules will be sufficient for grasp of AND.<sup>18</sup>

In a way, this supports Fodor’s (2004a, 2004b) position that having the concept AND (say) doesn’t depend on being disposed to accept its introduction and elimination rules, but rather on being able to think conjunctive thoughts (by having a concept that means *and*). I say ‘in a way’ because the considerations I have

---

<sup>16</sup> I want to be clear on an important point. I am not proposing here a *psychological* procedure to check the consistency of candidate Boolean functions with the introduction and elimination rules attached to a concept (although it is plausible that our deductive device is supplemented by a procedure to monitor for contradictions—see Sperber & Wilson 1995: 102). This would be to propose that our minds represent the truth tables for all Boolean functions, and can select the appropriate truth table on the basis of the inference rules attached to a concept. I do not make this claim. Rather, I am claiming that the introduction and elimination rules associated with a logical concept are metaphysically sufficient to specify the content of that concept.

<sup>17</sup> Cf. Gazdar & Pullum (1976), who discuss just such a constraint. I will not motivate the constraint here, because it is merely illustrative. I will demonstrate below, however, that a similar but more general constraint is well-motivated and psychologically real.

<sup>18</sup> Whether it is also sufficient for deploying conjunction in mental inference is a separate issue that I will return to below.

presented above show that accepting the introduction and elimination rules is not necessary for grasping the concept AND. This is consistent with Fodor's position that having AND doesn't depend on having these rules. But Fodor's position is more general. It's not that he thinks some of the canonical rules are unnecessary, it's rather that he denies that *accepting inference rules is constitutive of concept possession*. We will return to this point below.

If our minds have a constraint such as the one I have proposed, then the standard inferential role account of the content of AND is undermined. It is so far an open question, of course, whether our minds actually are like this, and whether the considerations set out above with respect to AND can be generalised to the other logical connectives. We also need to consider the question of whether a mind without AND-introduction could be effective at performing deductive inference. These are the questions I shall turn to in the following section.

### 3.5. Elimination rules and mental deduction

The reason that logicians have stated the meaning of 'and' in terms of the standard introduction and elimination rules is not just that these serve to fix the meaning of 'and' (that is, to uniquely specify the function it performs). After all, logicians can just specify the truth table to fix the meaning. The point about the introduction and elimination rules is that they are *rules*, and they are needed to support the process of deductive inference. The question then arises, if a mind dispensed with introduction rules, would it be able to perform deductive inferences?

It is not, in fact, implausible that our mental deduction device relies only on elimination rules, and no introduction rules. Sperber & Wilson (1995: Chapter 2) have argued convincingly that this is the case.<sup>19</sup> Given a set of premises, there is an

---

<sup>19</sup> See also §2.7 above. Cf. Gentzen (1935/1969: §5.13), who expressed the view that a logical constant was defined by its *introduction* rules, its elimination rules being just consequences of these rules (see Milne 1994; Tennant 1987: 94; Koslow 1992). Note that although interpretation may not invoke introduction rules, we presumably need something like introduction rules in production; I will

infinite set of conclusions that can be validly drawn using the standard introduction and elimination rules for the logical connectives. For example, from the assumption that  $p$  it is possible to derive an infinite number of conclusions of the form  $p \vee q$  (for any  $q$ , regardless of truth value) using the rule of *or*-introduction. As Sperber & Wilson point out, this is unproblematic in an informal system of natural deduction, where it is left to the intelligent user of the system to determine which rules to apply at which point in a derivation. But in any characterization of our mental deduction systems, such profligacy is highly problematic. There are two specific problems which Sperber & Wilson raise. The first problem is that in any formal model of our deductive device in which they are incorporated, introduction rules will apply an infinite number of times to any set of assumptions, generating an infinite set of conclusions. The second problem is that the conclusions they derive are trivial in a certain (intuitive) sense: conjoining any arbitrary proposition to an assumption through (for example) *or*-introduction does not produce a conclusion that is useful to an organism in the sense of improving its representation of (that is, its understanding of) the world.

There are two reasons why it might be considered necessary to postulate both introduction and elimination rules. The first reason is that although introduction rules directly derive only trivial conclusions, they appear to be necessary inasmuch as these trivial conclusions are themselves needed as premises for the subsequent derivation of non-trivial conclusions. The second reason is that introduction rules appear to be required as they are constitutive of the content of logical concepts. We will consider each of these in turn.

First, consider the need for introduction rules in derivations. In order to derive the conclusions in (5) and (7) below from the premises in (4) and (6), a step of *and*-introduction or *or*-introduction seems unavoidable (examples taken from Sperber & Wilson 1995: 98).

---

not speculate on this question here, other than to suggest that, however production is achieved, there is no particular reason to assume that it must call on meaning postulates attached to concepts.

- (4) a. If the trains are on strike and the car has broken down,  
there is no way of getting to work.
- b. The trains are on strike.
- c. The car has broken down.
- 

(5) There is no way of getting to work

- (6) a. If the boiler needs repairing or the electricity has been  
cut off, the house will be uninhabitable.
- b. The boiler needs repairing.
- 

(7) The house will be uninhabitable.

The reason for this is that a standard logical derivation would proceed as in (8) and (9) respectively.

- (8) a.  $(p \wedge q) \supset r$  [premise]  
b.  $p$  [premise]  
c.  $q$  [premise]  
d.  $(p \wedge q)$  [by  $\wedge$ -introduction from b. and c.]  
 $\therefore r$  [by modus ponens from a. and d.]

- (9) a.  $(p \vee q) \supset r$  [premise]  
b.  $p$  [premise]  
c.  $(p \vee q)$  [by  $\vee$ -introduction from b.]  
 $\therefore r$  [by modus ponens from a. and c.]

As can be seen, each of these derivations relies on the corresponding introduction rule. There is no reason to assume, however, that mental reasoning uses the standard rules of informal natural deduction. As Sperber & Wilson point out, in order to show that these introduction rules are required it would be necessary to show that the same derivations could not be carried out using alternative elimination rules, or that such rules were implausible on psychological grounds. But this is not the case. Sperber &



Wilson show that the conjunctive and disjunctive versions of modus ponens in (10) and (11) obviate the need for any introduction rules. What is more, they argue on theoretical grounds that such rules are psychologically plausible, and cite experimental evidence (from Rips 1983) in favour of (11).

(10) *Conjunctive modus ponens*

a. *Input:* (if ( $p$  and  $q$ ) then  $r$ )

$p$

*Output:* (if  $q$  then  $r$ )

b. *Input:* (if ( $p$  and  $q$ ) then  $r$ )

$q$

*Output:* (if  $p$  then  $r$ )

(11) *Disjunctive modus ponens*

a. *Input:* (if ( $p$  or  $q$ ) then  $r$ )

$p$

*Output:*  $r$

b. *Input:* (if ( $p$  or  $q$ ) then  $r$ )

$q$

*Output:*  $r$

I now turn to the second reason why it may be considered necessary to postulate introduction rules in mental logic, the fact that they have generally been taken to be constitutive of the content of the logical connectives. We have already seen above that the standard introduction and elimination rules governing ‘and’ are sufficient to uniquely specify the appropriate Boolean function. The same is true of ‘or’ and ‘if...then’, as I will now briefly show.

The canonical rules governing the use of ‘or’ are given in (12).

- (12) a.  $p / p * q$   
 b.  $q / p * q$   
 c.  $p * q, \neg p / q$   
 d.  $p * q, \neg q / p$

Together, these rules uniquely specify the corresponding Boolean function (function 7 in (13)). Rule a. eliminates functions with a value of '0' when  $p$  has the value '1'—that is, functions 0–2, 4–6, 8–10 and 12–14. Rule b. eliminates functions with a value of '0' when  $q$  has the value '1'—that is, functions 0–4, 6, 8–12, 14. Rules c. and d. both eliminate functions with a value of '1' when  $p$  has a value of '0' and  $q$  has a value of '0'—that is, functions 8–15.<sup>20</sup> The only function which has not been eliminated is function 7, the correct result.

(13)

		Or															
		If...then															
$p$	$q$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
0	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
1	0	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

As regards 'if...then', consider the introduction and elimination rules in (14) below.

- (14) a.  $\neg p / p * q$   
 b.  $q / p * q$   
 c.  $p * q, p / q$   
 d.  $p * q, \neg q / \neg p$

The introduction rules I have given in a. and b. are not the standard ones, although they do suffice to uniquely specify the appropriate function and logicians have claimed that "it is reasonable to require of any theory of the conditional that it

<sup>20</sup> The fact that both of these rules eliminate the same functions means that one can be considered redundant in constituting the content of OR. We will see below, however, that once introduction rules are eliminated, both elimination rules are necessary to fix the content.

explain why they hold”.<sup>21</sup> The standard introduction rule for ‘if...then’ is the conditional proof, which can be expressed in the following form: “if  $q$  can be derived from the assumption that  $p$ , by some sequence of rules of the deductive device, then it may be inferred that  $p \supset q$ , whether or not it is in fact the case that  $p$ ”. This rule cannot be stated in terms of the truth values of  $p$ ,  $q$  and  $p * q$ , and so cannot be used as a basis for eliminating candidate Boolean functions. In fact, however, the choice of introduction rules will not be important in what follows, since I propose that such rules can be eliminated in any case.

The rules in (14) specify the function corresponding to ‘if...then’ (function 13 in (13)) as follows. Rule a. eliminates functions with a value of ‘0’ when  $p$  has the value ‘0’—that is, functions 0–11. Rule b. eliminates functions with a value of ‘0’ when  $q$  has the value ‘1’—that is, functions 0–4, 6, 8–12 and 14. Rules c. and d. eliminate functions with a value of ‘1’ when  $p$  has the value ‘1’ and  $q$  has the value ‘0’—that is, functions 2, 3, 6, 7, 10, 11, 14, 15.<sup>22</sup> The only function which has not been eliminated is function 13, the correct result.

One additional consideration for OR and IF...THEN, which did not arise for AND, is that the introduction and elimination rules attached to these concepts involve negation, and therefore presuppose the availability of NOT. It is not clear what to do about NOT. Giving a suitable implicit definition or possession conditions for NOT is a difficult problem which has not been adequately solved (see Peacocke 2004). The standard rules of double negation introduction and elimination ( $\lceil p / \neg\neg p \rceil$  and  $\lceil \neg\neg p / p \rceil$ ), while they can be used to derive potentially useful rules such as modus tollens with a negated consequent ( $\lceil p \supset \neg q, q / \neg p \rceil$ ), cannot be employed to introduce or eliminate single instances of negation, nor to specify the truth table for

---

<sup>21</sup> See Martin (1987: 14). In fact, these two rules, although valid in propositional logic, have been regarded by many as counterintuitive as applied to the natural language expression ‘if...then’.

<sup>22</sup> Again, once introduction rules are eliminated we will see that both rule c. and rule d. are necessary for fixing the content of IF...THEN. (Cf. footnote 20 on p. 103 above.)

negation. In what follows I will assume that NOT is antecedently available to our minds—a reasonable, if unexplained, assumption.<sup>23</sup>

We have seen briefly how the introduction and elimination rules for the logical connectives ‘and’, ‘or’ and ‘if...then’ can fix the content of the corresponding concepts. We are now in a position to return to the question of whether it is possible for a mind that does not grasp the full set of canonical inference rules for a given logical connective to nevertheless possess the corresponding concept. We have already seen that given a certain general assumption about the mind, it is possible to grasp the concept AND by grasping only its elimination rules. I will argue in what follows that given a more general assumption about the mind (which I will show to be motivated), this argument can be extended to all of the standard logical connectives (conjunction, disjunction and implication).

The assumption that I want to argue for is that the meaning postulates (that is, inference rules) attached to a concept constrain the interpretation of that concept in two distinct ways, as set out in (15).<sup>24</sup>

---

<sup>23</sup> Since ‘ $\vee$ ’ and ‘ $\rightarrow$ ’ can be defined in terms of ‘ $\wedge$ ’ and ‘ $\neg$ ’, it follows that a mind which grasps the latter has all the resources necessary for grasp of the former. It does not, however, follow that such a mind has the concepts OR and IF...THEN. Possession of a logical concept is a question not of whether a mind has the necessary resources *in principle*, it is a question of whether a mind possesses a concept with the appropriate meaning (that is, in the case of logical concepts, whether it has a concept with the appropriate logical properties).

<sup>24</sup> Gazdar & Pullum (1976) propose a number of constraints that they employ to rule out certain connectives as lexicalizable in natural language. Although there are similarities in the general approach, the motivation, the particular constraints used, and the results, are very different. (Thanks to Jay Atlas for drawing my attention to the Gazdar & Pullum paper.)

(15) a. **Input consistency constraint**

For a concept  $C$  with attached meaning postulates  $MP_1, MP_2, \dots, MP_n$ , an interpretation of  $C$  that would make any of  $MP_1, MP_2, \dots, MP_n$  vacuous, in the sense that the input conditions for that postulate would be contradictory, is ruled out as a possible interpretation of  $C$ .

b. **Postulate validity constraint**

For a concept  $C$  with attached meaning postulates  $MP_1, MP_2, \dots, MP_n$ , an interpretation of  $C$  that would make any of  $MP_1, MP_2, \dots, MP_n$  invalid is ruled out as a possible interpretation of  $C$ .

These constraints, although they may have psychological justification (see below), are taken to be metaphysical rather than psychological in nature. That is, it's not that the mind *uses* the constraints to pick out the relevant truth table, and then reasons on the basis of this truth table. Rather, the mind reasons according to the meaning postulates attached to a concept, which has the (metaphysical) effect of ruling out some truth tables as interpretations of this concept.

A different way of stating the input consistency constraint is that any truth-table which makes the input conditions for any of the inference rules attached to a concept inconsistent is ruled out as a possible truth table for that concept. The justification I want to offer for the input consistency constraint is this. A meaning postulate whose input conditions were contradictory would have no utility, as there would be no logically possible circumstances under which it would apply. Not only would such a meaning postulate be useless, but its origin would also be obscure. A postulate could conceivably either be innate or acquired. But it is difficult to see how such a meaning postulate could be innate, given that it does not—nor could it ever have—served any purpose. It is similarly difficult to imagine how such a postulate could come to be acquired.

The postulate validity constraint is just a restatement of the status of meaning postulates: that is, any interpretation which would make the meaning postulates attached to a concept invalid, in the sense that the output could be false when the inputs were all true, is ruled out. As such, this constraint requires no further justification.

Given these two constraints, it can be shown that the contents of the concepts of conjunction, disjunction and implication are constituted by their elimination rules, without the need for introduction rules. A summary of the following results is provided in (20) below.<sup>25</sup>

Consider first the concept of conjunction, which has (at a minimum) the meaning postulates in (16).

(16) a. *Input:*  $p * q$

*Output:*  $p$

b. *Input:*  $p * q$

*Output:*  $q$

By the input consistency constraint, any truth table which makes the input conditions for any of the meaning postulates contradictory is ruled out. In this case, since each meaning postulate has only one input condition, any truth table on which that input is self-contradictory will be ruled out. This eliminates truth table 0.

By the postulate validity constraint, any truth table which makes one or more of the meaning postulates invalid is ruled out. This means that any truth table on which the input of one of the meaning postulates is true and the output false is ruled out. As we have already seen, this eliminates truth tables 2–15. The only remaining truth table is 1, the truth table for AND.

---

<sup>25</sup> Note also that these constraints can deal with problematic connectives such as ‘tonk’ (Prior 1960), defined by the inference rules  $\lceil p / p * q \rceil$  and  $\lceil p * q / q \rceil$ . Applying the constraints in (15), these inference rules eliminate all truth tables.

Next consider the concept of disjunction, which has minimally the meaning postulates in (17).

(17) a. *Input:*  $p * q$

$\neg p$

*Output:*  $q$

b. *Input:*  $p * q$

$\neg q$

*Output:*  $p$

The input consistency constraint rules out any truth table on which the two inputs of either meaning postulate are contradictory. This rules out truth tables 0–5. The postulate validity constraint rules out any truth table on which either of the meaning postulates is invalid, eliminating truth tables 8–15. This leaves two truth tables remaining (6 and 7), those for OR (inclusive ‘or’) and XOR (exclusive ‘or’), an interesting result.

A grasp of the introduction and elimination rules for ‘or’, as given in (12), is sufficient for grasp of OR, but excludes XOR, as we saw earlier (see the table in (13)). That is, the full set of canonical inference rules uniquely specifies inclusive ‘or’. A grasp of the elimination rules only, together with the general principles I have proposed, can eliminate all non-disjunctive possibilities, but leaves open both an OR and XOR interpretation. Let us call this underspecified concept DISJUNCTION. This result is interesting, because it has often been claimed that English ‘or’ can have both inclusive and exclusive interpretations.<sup>26</sup> On the current proposal, both of these interpretations are left open.

---

<sup>26</sup> There has been fairly extensive discussion of this question in the literature. No clear consensus has been reached, and the three obvious possibilities have all been argued for: that English ‘or’ is always inclusive (Pelletier 1977, Lepore 2000), that it is (virtually) always exclusive (Lakoff 1971, van Dijk 1977: 63), and that ‘or’ has two possible meanings, the one to be adopted being determined by

This suggests one of two possibilities. The first is that DISJUNCTION could be genuinely ambiguous at the semantic level between inclusive and exclusive readings. There is no strong reason to believe that this is the case, however. On the assumption that English ‘or’ expresses the concept DISJUNCTION, the ‘ambiguity hypothesis’ would lead us to expect that sentences with ‘or’ would be regarded by speakers as ambiguous in the same way that sentences with polysemous lexical items are, which is not the case. The second possibility is that DISJUNCTION expresses a general, underspecified,<sup>27</sup> meaning, which can be further specified through pragmatic processes such as concept narrowing. This is the much more likely possibility. For example, we could suppose that a pragmatic principle assigns the more general meaning—that is, the one giving rise to fewer entailments, in this case OR—in situations such as this (cf. Chierchia & McConnell-Ginet 2000: 78–79) and that the narrower meaning—in this case, XOR—could arise via conversational implicature (specifically, a scalar implicature). See Grice (1967), Sperber & Wilson (1995: Postface), Carston (1998), Noveck (2004), and Crain et al. (2005).

It could also be the case that additional meaning postulates are the mechanism through which concept narrowing is achieved. For example, additional elimination rules such as those expressed by the meaning postulates  $\lceil p * q, p / \neg q \rceil$  and  $\lceil p * q, q / \neg p \rceil$  would act in this way, as they serve to rule out the inclusive reading and therefore specify the (narrower) concept XOR. This could be a way to account for languages such as Latin, which had separate lexical items ‘vel’ and ‘aut’, according to some accounts corresponding to inclusive and exclusive disjunction, respectively.<sup>28</sup> Such an account could also be adopted if English ‘or’ can be

---

pragmatic factors (Hurford 1974). See Evans et al. (1993: Chapter 5) and Jennings (1994: Chapter 3) for reviews.

<sup>27</sup> Note that to say that the concept is *underspecified* is not to say that it is *indeterminate*. This distinction is discussed in more detail below in relation to implication.

<sup>28</sup> This analysis, although the received wisdom in most introductory logic texts, is controversial. See Jennings (1994: Chapter 9) for convincing arguments that Latin ‘vel’ and ‘aut’ should not be seen as corresponding to inclusive and exclusive disjunction. Gazdar & Pullum (1976) also express their doubts.



demonstrated to have only an exclusive interpretation.<sup>29</sup> Notice, however, that no additional set of meaning postulates based on elimination rules can perform the opposite function—that of eliminating the narrower meaning (XOR) in favour of the more general meaning (OR). It follows that if meaning postulates are restricted to elimination rules, as we have been supposing, the concept OR cannot be uniquely specified (other than by pragmatic means).<sup>30</sup>

Finally, consider the concept of implication, which has at least the meaning postulates in (18).<sup>31</sup>

---

<sup>29</sup> Cf. footnote 26 (p. 108) above.

<sup>30</sup> The following is an informal proof of this. Suppose that some set of meaning postulates based on elimination rules could rule out XOR. Now, on the approach we are taking there are two possible ways to rule out a truth table: the input consistency constraint, and the postulate validity constraint. Consider each in turn. First, it can be shown that the input consistency constraint cannot rule out XOR. The reason is that there is no condition where XOR is false when any one of  $p$ ,  $q$ ,  $\neg p$ ,  $\neg q$  are true, and therefore no possibility for a postulate with inconsistent inputs. Next, it can be shown that the postulate validity constraint cannot rule out truth tables satisfying the condition  $\langle |p| = 1, |q| = 1, |p * q| = 0 \rangle$  (the only condition on which OR and XOR differ), because this constraint works by ruling out truth tables that make the output of the postulate false when its inputs are all true. But, on the assumption there are no introduction rules, the string ' $p * q$ ' must be an input to any postulate, implying that this constraint only looks at conditions where  $|p * q| = 1$ , which it is not in the case we are considering. By analogous reasoning, it is not possible (other than by pragmatic means) to uniquely specify the concepts CONDITIONAL, REVERSE-CONDITIONAL or NAND via elimination rules. A systematic analysis of the full range of meaning postulates and connectives is provided in the appendix to this chapter.

<sup>31</sup> The psychological evidence suggests, in fact, that while the first of these meaning postulates, corresponding to modus ponens, is directly represented and highly accessible, the second meaning postulate, corresponding to modus tollens, is not directly represented, or at least not highly accessible (see Evans et al. 1993: Chapter 2). The effects of modus tollens, and hence the metaphysical constraints it places on possible truth-functional interpretations, can be obtained from modus ponens together with a form of reductio ad absurdum: given the premises of modus tollens ( $p * q, \neg q$ ) assume that  $p$ , from which by modus ponens we can conclude that  $q$ , contradicting one of the initial premises and falsifying our assumption that  $p$  (see Evans et al. 1993: 14–15 and Braine & O'Brien 1991 for discussion and detailed proposals; for psychological evidence in support of the proposal that in evaluating conditionals people create an imaginary world that includes the assumption that  $p$ , or at least focus on the possibility that the antecedent is true, see Hadjichristidis et al. 2001, Over & Evans 2003 and Evans et al. 2003, as well as Sperber et al. 1995). Alternatively, it cannot of course be ruled out that the only meaning postulate in this case is modus ponens, and that there are no other metaphysical constraints in play. This would give rise to an underspecified concept allowing AND, Q-IDENTITY, CONDITIONAL and BICONDITIONAL interpretations. Perhaps Q-IDENTITY can be generally excluded on pragmatic grounds: why use a connective when the antecedent is always irrelevant to the

(18) a. *Input:*  $p * q$

$p$

*Output:*  $q$

b. *Input:*  $p * q$

$\neg q$

*Output:*  $\neg p$

The input consistency constraint for these two meaning postulates rules out truth tables 0, 1, 4, 5, 8 and 12. The postulate validity constraint rules out truth tables 2, 3, 6, 7, 10, 11, 14 and 15. As with disjunction, this leaves two truth tables remaining—this time, 9 and 13, the truth tables for the biconditional and the conditional.

This is also an interesting result. The difference between the conditional and the biconditional is that the former, but not the latter, is true when the antecedent is false and the consequent is true. This is precisely the condition that people have most intuitive difficulty with.<sup>32</sup> Indeed, natural language utterances containing ‘if...then’ can often have both conditional and biconditional interpretations.<sup>33</sup> Consider for

---

truth value, rather than just stating  $q$  (cf. Gazdar & Pullum 1976)? The remaining three possible interpretations could be further specified pragmatically (by narrowing, assuming a general principle assigning the most general interpretation). Interestingly, this might explain the characteristic pattern of errors associated with conditional reasoning, including those of children, who seemingly go through three developmental stages: they initially interpret conditional statements as existential conjunctions, later as biconditionals, and finally as true conditionals (see Evans et al. 2003, Barrouillet & Lecas 1998, and Barrouillet et al. 2000). Sperber et al. (1995) show how pragmatic considerations can give rise to conjunctive interpretations for conditional statements.

<sup>32</sup> Although there are several possible reasons for this, as the natural language uses of ‘if...then’ are not limited to expressing truth-functional relations.

<sup>33</sup> The situation is similar to that of disjunction (see footnote 26 on p. 108 above), and the corresponding literature on implication is vast. Some consider that English ‘if...then’ expresses the conditional (material implication) and never the biconditional (material equivalence). Others allow that it may also (and perhaps always in young children) express the biconditional, which can explain why subjects often draw the ‘fallacious’ conditional inferences—denial of the antecedent and affirmation of the consequent. Many, however, consider that it has a significant non-truth-conditional component. See Evans et al. (1993: Chapter 2) for a review.

example the utterance in (19).<sup>34</sup>

(19) If the train is on time then I'll be at your office at four o'clock

This utterance can be interpreted on either a conditional or a biconditional sense of 'if...then'. It could be interpreted as stating merely a *sufficient* condition (the train being on time) for arriving at four o'clock, in which case 'if...then' is to be given a conditional interpretation (even if the train is late, I could rush to the office and still be there at four o'clock). Alternatively, the utterance could be interpreted as stating a *necessary* condition for arriving at four o'clock (I'll be rushing already, so if the train is late I certainly won't be at the office at four o'clock).

As with disjunction, then, the present analysis captures the range of observed truth-functional interpretations. Again, there are two possibilities: semantic ambiguity or underspecification. For the same reasons as previously discussed, the latter seems preferable. Note that to say the concept is *underspecified* is not to say that some of its truth conditions are *indeterminate*. To see this, compare a classical two-valued logic of the kind assumed here with a three-valued logic allowing the values true, false and indeterminate. In the system proposed here, all cells in a truth table must, when filled at all, be filled either with a '0' or a '1'. However, some truth tables are underspecified in the sense that the metaphysical constraints imposed by the meaning postulates attached to a concept may not be sufficient to determine the values of these cells, which may be done by pragmatic concept-narrowing processes. By contrast, it has been suggested by Wason (1966), Johnson-Laird & Tagart (1969), Evans & Over (2004) and others that certain truth tables might be defective, in the sense that the values of certain cells could be indeterminate (particularly the cells corresponding to a false antecedent in the truth table for the conditional) because a conditional statement is irrelevant in such circumstances (see Evans et al. 1993, Chapter 2). This is to propose a determinate three-valued logic, since there is no lack

---

<sup>34</sup> Cf. Partee et al. (1993: 102–104).

of specification here—the indeterminate cells are not awaiting a value, they already have one (viz., ‘indeterminate’, in the sense of irrelevant). This point is important, because the ‘defective truth table’ account has certain highly implausible implications (see Johnson-Laird 2005).

I’ll refer to the underspecified concept as IMPLICATION. Additional meaning postulates or pragmatic concept narrowing could provide further specification of IMPLICATION. Pragmatic narrowing would work analogously to the case of DISJUNCTION—again, we have a more general meaning (CONDITIONAL) and a narrower meaning (BICONDITIONAL), and the latter could arise as an interpretation of the former via scalar implicature (see above). Alternatively, additional elimination rules such as those expressed by the meaning postulates  $\lceil p * q, q / p \rceil$  or  $\lceil p * q, \neg p / \neg q \rceil$  would serve to rule out the material implication reading and therefore specify the (narrower) concept BICONDITIONAL. As with DISJUNCTION, no set of meaning postulates based on elimination rules could rule out the BICONDITIONAL meaning.<sup>35</sup>

All the above results for conjunction, disjunction and implication are summarized in the table in (20) below.<sup>36</sup>

---

<sup>35</sup> Cf. footnote 30 (p. 110) above.

<sup>36</sup> In this table, ‘ICC’ stands for the input consistency constraint, and ‘PVC’ stands for the postulate validity constraint.

(20)

		Contradiction	And	p and not q	p-identity	q and not p	q-identity	Xor	Or	Nor	Biconditional	Not-q	Reverse-conditional	Not-p	Conditional	Nand	Tautology
p	q	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
0	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
1	0	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

Conjunction	ICC	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	PVC	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Disjunction	ICC	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	PVC	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Implication	ICC	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	PVC	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

It has been argued by Sperber & Wilson (1995) that our mental deductive device plausibly operates with elimination rules but no introduction rules. I have argued above, on independent grounds, that given a reasonable assumption about the mind the content of the logical connectives can be determined (with a certain degree of principled underspecification) purely on the basis of the proposed meaning postulates (which correspond to the canonical elimination rules for these connectives). This undermines inferential role accounts of logical content, according to which grasping a logical concept *just is* to grasp both its introduction and elimination rules. In doing so, the present account also provides support for Fodor's (2004a, 2004b) claim that possessing a logical term is a matter not of being disposed to accept the term's canonical inferences, but of possessing a concept with the right meaning. His claim, then, is more general, as he does not believe that *any* meaning postulates are constitutive of logical content.

In the following section I address the question of whether meaning postulates are to be seen as content constitutive.

### 3.6. Are meaning postulates content constitutive?

What can we now say about whether meaning postulates are content constitutive? First, note that the discussion above is at least suggestive of meaning postulates being constitutive of the content of logical terms. After all, we have seen how meaning postulates impose metaphysical constraints on the content of the concept to which they are attached, which would tend to indicate that meaning postulates are content constitutive. In this section, we will consider whether such a claim can be supported.

In deciding on the constitutivity of meaning postulates, two tests are important:

1. Can we have cases where we possess a logical concept *without* having the associated meaning postulates? and
2. Can we have cases where we possess a logical concept with the associated meaning postulates, but where these meaning postulates *fail* to make a contribution to content?

I will discuss each of these tests in turn.

#### *Content without meaning postulates?*

The first test is important because if it can be shown that the content of a logical concept may be constituted in the absence of any prescribed set of meaning postulates, this could suggest that meaning postulates are not content constitutive. After all, if an inference rule is not necessary for the possession of a concept, it is difficult to see how that inference rule could be constitutive of the content of that concept.

Inferential role accounts of content, for example, deny that we can possess a logical concept without possessing/grasping the canonical inference rules (that is, meaning postulates) for that concept. According to inferential role accounts, the identity conditions for a concept are the same as the possession conditions for that concept. It follows on such an account that if a given inference rule is not among the

possession conditions for a concept, then that inference rule cannot be content constitutive. Or, to put things the other way round, on an inferential role account any meaning postulate which is content constitutive is also required for possession of the concept.

In fact, however, it is clear from the earlier discussion that content can be constituted without the need for any particular set of meaning postulates. To see this, first note that there is no a priori reason why we need meaning postulates at all in order to fix the content of a logical connective (although we may of course need them for other reasons, such as for mental deduction). There is nothing incoherent about the idea of a mind which represents logical properties in some other way—say, by directly representing the relevant truth table.

Beyond this, we have seen that different combinations of meaning postulates and general principles can serve to pick out the same truth table (see also the appendix to this chapter). For example, a concept ‘\*’ will be the concept AND if it has attached meaning postulates corresponding to the canonical introduction and elimination rules for conjunction, given in (2) above. It is clear, however, that these three meaning postulates are not *required* in order to fix the content of AND. We have already seen above that, given a certain plausible assumption about the mind, we can dispense with the meaning postulate corresponding to ‘and’-introduction. Even restricting ourselves to elimination rules, there can be different sets of meaning postulates that fix the content of a particular connective.<sup>37</sup>

Given that there is no particular meaning postulate or set of meaning postulates that is required in order to fix the content of a connective, must we conclude that meaning postulates are not content constitutive? Perhaps not. The other possibility is to allow that, *contra* inferential role theorists, something can be constitutive of the content of a concept without being *necessarily* present (that is, without being a

---

<sup>37</sup> Of course, to say that there are different sets of meaning postulates that would serve to pick out the same logical content is not intended to make any commitments as to whether, in point of psychological fact, the full set of possible meaning postulates is actually available to us.

possession condition). I will argue in favour of such a proposal, and against the inferential role account.

The position taken by inferential role theorists is influenced by the fact that in cases such as conjunction there do not seem to be any alternatives: if not the canonical introduction and elimination rules, then what other rules could determine the content of AND? If another non-equivalent set of rules could be found which could also determine the content of AND, then the identification of content-constitutive inferences with possession conditions would be undermined. So, to allow that a meaning postulate can be content constitutive without necessarily being present is just to allow that there can be more than one way to constitute the content of a concept.

In fact, of course, the previous section demonstrated just that: different combinations of meaning postulates and general principles can serve to fix the content of a logical connective (further discussion is provided in the appendix to this chapter). In which case, there is no justification for identifying content-constitutive inferences with possession conditions. All that we can say of meaning postulates is this: *if* a particular meaning postulate is present, then it constitutes (partly or wholly) the content of the concept. But this is not to say that such a meaning postulate *must* be present.

There is a potential objection to this line of reasoning, which is that although meaning postulates may play a critical role in fixing content, this does not mean that they have to *constitute* content. After all, it is perhaps not *prima facie* implausible to see meaning postulates as just another kind of sustaining mechanism. Compare: a mechanism to recognize tigers may play a critical role in fixing TIGER-content, but this doesn't mean that such a mechanism is constitutive of TIGER-content. On such a view, one could regard the abstract logical properties of the concept (that is, its truth table) as giving the content, and the meaning postulates as merely providing a means of what Fodor would call "semantic access" to the content. It's *that* the right logical properties are picked out, not *how* they're picked out, that's important. In the same



way, a tiger-detector provides a sustaining mechanism linking TIGER with its content, the abstract property of *tigerhood*. It's *that* the concept picks out the right property, not *how* it does so, that's important to informational semantics.

On reflection, however, it's not clear that this will do. According to an informational semantic account, the concept TIGER (and the concept BACHELOR, and even the concept DOORKNOB) means what it does because of a nomological link between the concept and the corresponding property. As such, it's the existence of the link, not the mechanisms by which the link is sustained, that determine the content. But the situation is different for the logical terms. It would be distinctly odd to claim that it is in virtue of some nomological link between AND and the property *conjunction* that AND means what it does.<sup>38</sup> It is not plausible to see AND and other logical concepts as *referring* to some abstract logical property. Rather, AND means what it does because it possesses the logical properties of conjunction—a token of AND is an *instance* of conjunction.<sup>39</sup> Compare this with TIGER, which refers to tigers, but which doesn't have the properties of tigers (unlike tigers, TIGERS have no stripes)—a TIGER-token is certainly not an *instance* of tigerhood. The right question to be asking about AND, therefore, is not how it gets linked to the property it expresses, but what gives it the properties that it possesses. Sustaining mechanisms can't provide an answer to this question, but meaning postulates just might.

I have argued that there is no privileged set of content-constitutive meaning postulates, but that this does not mean that meaning postulates are not content constitutive. Rather, meaning postulates are content constitutive where present, but no particular meaning postulate is required to be present. Should you find this argument unconvincing, however, the second test proves to be conclusive.

---

<sup>38</sup> Cf. Prinz & Clark (2004) and Fodor's (2004b) reply.

<sup>39</sup> Assuming the usual caveats concerning cases where AND is mentioned rather than used.

*Meaning postulates without content?*

The second test is in fact the crucial one for determining whether meaning postulates are content constitutive. If meaning postulates can be present without being (partly or wholly) constitutive of content, then meaning postulates clearly cannot be content constitutive.

Suppose for the sake of argument that meaning postulates could be present without being constitutive of content. This would be to claim that although the mind might make use of meaning postulates for reasoning, they do not fix the content of the concepts to which they are attached. But it is difficult to see how this could be the case. Suppose that a mind has a concept ‘\*’ with the meaning postulates  $\lceil p * q, \neg p / q \rceil$  and  $\lceil p * q, \neg q / p \rceil$  attached. These meaning postulates serve to pick out truth tables 6 and 7 in (20), given the constraints set out in (15). Suppose now that the mind acquires a new set of inference rules,  $\lceil p * q, p / \neg q \rceil$  and  $\lceil p * q, q / \neg p \rceil$ , which it attaches to ‘\*’ as meaning postulates. The crucial point to note is that the effect of attaching these meaning postulates is to *thereby* change (/narrow) the content of ‘\*’ from DISJUNCTION to XOR. For now the set of meaning postulates attached to ‘\*’ picks out a different set of truth tables—that is, just truth table 6. And it just cannot be the case that the content of a connective is different from the truth table with which it is associated.

This same example, however, also raises a potential difficulty. For notice that in order to narrow the content of ‘\*’ from DISJUNCTION to XOR we have proposed the addition of two further meaning postulates,  $\lceil p * q, p / \neg q \rceil$  and  $\lceil p * q, q / \neg p \rceil$ . While it is reasonable to assume the addition of both of these postulates, in order that our mental deduction can be systematic, either one of these postulates alone is sufficient to narrow the content of ‘\*’ to XOR.<sup>40</sup> Since one of these postulates will be unnecessary for specifying the content, does this mean that only one is content constitutive? Which one?

---

<sup>40</sup> See the appendix to this chapter for a systematic analysis.

In fact, this seems to be the wrong way of looking at things. Whether a meaning postulate is content constitutive or not is a question of whether it constrains the (semantic) interpretation of the concept or not. It is irrelevant whether the constraints it imposes are spurious—what determines whether a postulate is content constitutive is whether it has an *input* into content, not whether that input has any *effect* on the content in a particular case. I would therefore argue that the right thing to say in a situation such as the one we have been considering is that both meaning postulates are content constitutive, and that their contribution to content (that is, the constraints they impose) happens to be identical. If either postulate was removed, the constraints imposed by the other would be substantive.<sup>41</sup>

We can see from the above considerations that, when present, meaning postulates cannot fail to be constitutive of content. This I take to demonstrate conclusively that the meaning postulates attached to a concept are content constitutive.

Note, by the way, that the current account of logical content avoids Fodor's (2004a) criticism that inferential role accounts are viciously circular (discussed in §2.4.4 above). Fodor's point was that if content is constituted by possession conditions, then it is circular to formulate the possession condition for a concept by employing that same concept. Fodor's argument was that any formulation of the possession conditions for AND, say, would need to employ the concept of conjunction. In the present account, a clear distinction has been made between possession conditions and content-constitutive inference rules. In the formulation I have given, the meaning postulates attached to a concept do not need to make use of the concept that they are attached to, except in an innocuous way. That is, they employ a conceptual placeholder '\*', which can be read as something like "this

---

<sup>41</sup> Compare this with inference rules that are not represented as meaning postulates. We can imagine a situation, say, where we have explicitly learned the inference rule *modus tollens*, but can only make use of this rule in conscious, reflective thought. That is, we have the rule, but not in the form of a meaning postulate. Such a rule *would* have a substantive effect on the content of the intuitive concept if it were in the form of a meaning postulate, but since it is not represented in the right way, it has no input into content—it is not content constitutive.

concept”. This is a standard feature of elimination rules, and there is no danger of circularity. For example, a meaning postulate  $\lceil p * q / p \rceil$  attached to a concept indicates how the concept to which it is attached may be eliminated from an expression of the right form. Such a meaning postulate operates in a purely syntactic manner, and is not sensitive to the meaning of ‘\*’.

#### *Meaning postulates and the non-logical vocabulary*

We have seen from the discussion above that meaning postulates, when present, are content constitutive. This does not mean, however, that they must be present, as there can be more than one way to constitute the content of a given concept. The reason that such a possibility is not normally considered is that it has been widely assumed (particularly by inferential role theorists) that content-constitutive inference rules and possession conditions are identical. Once this assumption is rejected, as we have seen that it must be, there is no reason to insist that content-constitutive inference rules must be present.

In chapter 2, it was argued that no principled distinction can be drawn between the logical and non-logical vocabularies. In the present chapter it was argued that content-constitutive inference rules (meaning postulates) are what account for the content of the logical vocabulary. It follows, by extension, that meaning postulates account for the logical properties of words which do not have purely logical content. For example, ‘but’ has the same logical properties as ‘and’, so it would be reasonable to assume that the logical content of ‘but’ is accounted for in the same way as the content of ‘and’—via content-constitutive meaning postulates. In this case, however, such meaning postulates do not exhaust the content. In addition to its logical properties, ‘but’ also introduces some non-truth-conditional notion of ‘denial of expectation’ or ‘contrast’ (see §2.3 above). In chapter 2, I proposed that we should make a distinction between logical and non-logical *content*, rather than between logical and non-logical *vocabularies*. Logical content is to be accounted for (inter

alia) via meaning postulates,<sup>42</sup> while non-logical content can be accounted for via informational atomism.

I have also presented arguments in chapter 2 that meaning postulates attached to a range of concepts that have generally been regarded as non-logical can account for our spontaneous inferential capacities. For example, we might propose that attached to our concept TIGER is a meaning postulate such as  $\lceil \varphi \text{ TIGER } \psi \rightarrow \varphi \text{ ANIMAL } \psi \rceil$ . If indeed there are such meaning postulates, the question arises whether these too are content constitutive. In the case of logical connectives, it was clear that some account of logical content was needed. Informational semantics could not provide such an account, and meaning postulates were an obvious alternative. In the case of predicative concepts (that is, concepts which express properties: natural kind concepts, nominal kind concepts, artefact concepts, and so on) informational semantics gives an account of the content. Are meaning postulates such as  $\lceil \varphi \text{ TIGER } \psi \rightarrow \varphi \text{ ANIMAL } \psi \rceil$  also partly constitutive of content?

It seems that they must be.<sup>43</sup> For if the meaning postulates attached to logical connectives are content constitutive, and if no principled distinction can be drawn between logical concepts and the rest, then we cannot avoid the conclusion that meaning postulates in general are content constitutive. And this seems right. If we have a meaning postulate such as  $\lceil \varphi \text{ TIGER } \psi \rightarrow \varphi \text{ ANIMAL } \psi \rceil$  attached to our concept TIGER, then ipso facto we cannot help thinking of tigers as animals (even if we were to find out that they are not). That is, the meaning postulate constrains our interpretation of the concept, hence constrains the content of the concept.

---

<sup>42</sup> I say ‘inter alia’, because there may be other mechanisms apart from meaning postulates that account for some aspects of logical content. For example, it may not be possible to fix the content of NOT via meaning postulates (see Peacocke 2004). There are also questions about how to account for the content of operators (such as ‘possibility’ and ‘necessity’) and quantifiers. In addition, as I have mentioned earlier, we probably need to think in terms of a broader distinction between *procedural* and *referential* meaning, with meaning postulates being only one aspect of procedural meaning (see Blakemore 2000; Carston 2002: §2.3.7).

<sup>43</sup> Cf. de Almeida (1999) who makes precisely the opposite claim that while meaning postulates are what account for conceptual connectedness, they are not content constitutive.

Of course, it is possible to imagine a concept TIGER', whose informational-semantic content is identical to that of the concept TIGER, but which does not have any meaning postulates attached. This concept TIGER' also expresses the property *tigerhood*, to which it is nomologically linked. The content of the two concepts is different, however, since the two concepts play a different role in mental deduction: TIGER, but not TIGER', is constrained in its interpretation to referring to a kind of animal *and this is so even if it turns out that tigers are not animals*. That is, even though meaning postulates are content constitutive, they are not guaranteed to be veridical.

One of the claims of this thesis is that our intuitive inferential capacities are mediated by meaning postulates attached to concepts, and that this is the case not just for typical logical concepts, but for many (if not most) of our concepts. However, there are many concepts that do not have a role to play in intuitive inference. These concepts can only be deployed in deliberate, reflective thought. Chapter 4 will explore in detail this important distinction between intuitive and reflective concepts.

## Appendix: The range of available meaning postulates

In what follows, I will give a more systematic analysis of the range of meaning postulates available, and how these can be combined in order to specify particular Boolean connectives. First, note that there are a number of limitations on the form that meaning postulates based on elimination rules can take:

1. All such postulates must have  $\lceil p * q \rceil$  as one of their inputs (they would not otherwise be eliminating ‘\*’).
2. No such postulate can have  $\lceil p * q \rceil$  as its output (again, such a postulate would not be eliminating ‘\*’).
3. No such postulate can have as its output one of its inputs (such a postulate would not derive any new information) or the negation of one of its inputs (such a postulate would be contradictory).

These limitations allow only five possible inputs (1–5 in the table below) and four possible outputs (a.–d. in the table below), which together give rise to the following 12 possible meaning postulates.<sup>44</sup>

---

<sup>44</sup> I assume in what follows that only two propositional variables are involved. We have seen that it may be necessary to propose meaning postulates involving three variables in order to provide a basis for certain mental inferences in the absence of introduction rules. For example, meaning postulates for conjunctive and disjunctive modus ponens were discussed in §3.5 above, and these incorporated three propositional variables. This raises the question of whether meaning postulates incorporating three (or more) propositional variables could be used to uniquely specify the connectives ‘or’, ‘nand’, ‘conditional’ and ‘reverse-conditional’, something which is not possible via meaning postulates incorporating two variables. I am inclined to believe that introducing additional propositional variables will not help in this regard, although I do not provide any proof of this here.

		Meaning postulate	T-tables ruled out by input consistency constraint	T-tables ruled out by postulate validity constraint	T-tables not ruled out
1.	a.	$p * q / p$	0	4-15	1-3
	b.	$p * q / q$	0	2-3, 6-15	1, 4-5
	c.	$p * q / \neg p$	0	1-3, 5-7, 9-11, 13-15	4, 8, 12
	d.	$p * q / \neg q$	0	1, 3-7, 9, 11-15	2, 8, 10
2.	b.	$p * q, p / q$	0, 4, 8, 12	2-3, 6-7, 10-11, 14-15	1, 5, 9, 13
	d.	$p * q, p / \neg q$	0, 4, 8, 12	1, 3, 5, 7, 9, 11, 13, 15	2, 6, 10, 14
3.	b.	$p * q, \neg p / q$	0-3	8-15	4-7
	d.	$p * q, \neg p / \neg q$	0-3	4-7, 12-15	8-11
4.	a.	$p * q, q / p$	0, 2, 8, 10	4-7, 12-15	1, 3, 9, 11
	c.	$p * q, q / \neg p$	0, 2, 8, 10	1, 3, 5, 7, 9, 11, 13, 15	4, 6, 12, 14
5.	a.	$p * q, \neg q / p$	0-1, 4-5	8-15	2-3, 6-7
	c.	$p * q, \neg q / \neg p$	0-1, 4-5	2-3, 6-7, 10-11, 14-15	8-9, 12-13

From this it is possible to determine which of the 16 Boolean connectives can be uniquely specified—that is, for which connectives there exist combinations of meaning postulates that can eliminate all truth tables except the desired one.<sup>45</sup> As shown in the table below, this is possible for all connectives except contradictions and tautologies (which are ruled out by definition) and ‘or’, ‘nand’, ‘conditional’ and ‘reverse-conditional’ (which cannot be uniquely specified to the exclusion of their more specific counterparts—‘xor’ for ‘or’ and ‘nand’, and ‘biconditional’ for ‘conditional’ and ‘reverse-conditional’):

---

<sup>45</sup> Of course, some of these connectives may not have any utility, for example because the truth value of the antecedent or the consequent is irrelevant to determining the truth value of the whole. For detailed discussion, see Gazdar & Pullum (1976).



Truth table	Meaning postulates required <sup>46</sup>
0. (contradiction)	— (by the input consistency constraint, no meaning postulate can be contradictory)
1. (and)	1a. + 1b.
2. ( <i>p</i> and not <i>q</i> )	1a. + 1d.
3. ( <i>p</i> -identity)	4a. + 5a.
4. ( <i>q</i> and not <i>p</i> )	1b. + 1c.
5. ( <i>q</i> -identity)	2b. + 3b.
6. (xor)	2d. + 3b. + 4c. + 5a.
7. (or)	— (3b. + 5a. gives OR/XOR)
8. (nor)	1c. + 1d.
9. (biconditional)	2b. + 3d. + 4a. + 5c.
10. (not- <i>q</i> )	2d. + 3d.
11. (reverse-conditional)	— (3d. + 4a. gives REVERSE-CONDITIONAL/BICONDITIONAL)
12. (not- <i>p</i> )	4c. + 5c.
13. (conditional)	— (2b. + 5c. gives CONDITIONAL/BICONDITIONAL)
14. (nand)	— (2d. + 4c. gives NAND/XOR)
15. (tautology)	— (by assumption, there are no tautological meaning postulates)

The four functions which cannot be uniquely specified via meaning postulates could, however, be specified in other ways. As discussed in the main text, a (pragmatic) principle could be postulated which directs us to the more general meaning in such situations, with the other meaning arising through conversational (scalar) implicature. See Grice (1967), Sperber & Wilson (1995: Postface), Carston (1998), Chierchia & McConnell-Ginet (2000: 78–79), and Noveck (2004).

---

<sup>46</sup> In some cases, more than one set of meaning postulates can be employed to uniquely specify a connective. For example, ‘and’ can be specified by [1a. + 1b.], but also by [1a. + 2b.]. In the table, I have listed only the (intuitively) most basic set. In other cases (‘xor’ and ‘biconditional’), some of the 4 meaning postulates listed are not strictly required (any three of the four would suffice, as well as some pairs). However, I consider ‘xor’ and ‘biconditional’ to be further specifications of ‘or’ and ‘conditional’, respectively, and hence to have all of the meaning postulates associated with their more general counterpart. It is then appealing (from a psychological viewpoint) to propose an additional two postulates, rather than just one, since these are symmetrical, and it seems plausible that if one of them is present then the other will be too. (In essence, I am doubting that our mental logic is non-systematic, in the sense that we could deduce  $p \vee q, p \vdash \neg q$ , say, but not  $p \vee q, q \vdash \neg p$ .)

## 4. Intuitive and reflective concepts

### 4.1. The problem of incomplete concepts

#### 4.1.1. Overview of the problem

In §1.3.2 above, we briefly considered the following kind of difficulty for Fodor's account of concept possession. Keil & Wilson (2000: 316) ask us to suppose that a biologist discovers the existence of a long-extinct mammal (a 'shmoo') after finding a unique mammalian genetic fragment in amber. The fragment is just enough to determine that it was from a mammal, different from all other known mammals, but not enough to determine the kind of mammal. Keil & Wilson claim that based on this information we can possess a concept SCHMOO, but that it cannot be the case that we are locked to the corresponding property (*shmoo*), since we don't know what that property is.

If we accept the relevant intuitions, then this case is—at least *prima facie*—problematic for Fodor's account of concept possession. Recall that, for Fodor, to possess a concept is to be locked to the property expressed by that concept. The present case, however, seems to be one where we possess a concept without being locked to the corresponding property. This suggests that locking is not necessary for concept possession, which would undermine Fodor's account.

In fact, there is a whole range of cases where we plausibly possess a concept without apparently being locked to the corresponding property—because we do not happen to have encountered instantiations of the property, because the property is not in fact instantiated or perhaps never was instantiated, or because there is no such property (in the case of nomologically or even logically impossible properties). Let us consider some examples, before discussing in more detail the difficulties that they raise:

#### ECHIDNA

Suppose that you have never encountered an echidna (or even pictures or other representations of echidnas), but that you have some rudimentary information about echidnas (that they are a kind of animal living in Australia, say). The property is instantiated, but you have never had the relevant experiences.

#### DODO

Since there are no longer any dodos, you can never have encountered one. But suppose that you have seen drawings of dodos. Although the property is not (currently) instantiated, you have nevertheless had relevant dodo-experiences (that is, experiences of dodo-representations).

#### SCHMOO

In the case suggested by Keil & Wilson, the property *schmoo* is not (currently) instantiated, and neither is it possible that you have had any experiences of schmoo-representations (since there are none).

#### UNICORN

Let's grant that unicorns are nomologically possible creatures.<sup>1</sup> The property *unicorn* is not (and never has been, and presumably never will be) instantiated. There are, however, unicorn-representations (pictures in story-books and so forth), so you may have had relevant experiences.

#### GHOST

Ghosts are, let us suppose, nomologically impossible (for example, they occupy a location in space but can pass through solid objects).<sup>2</sup> The property *ghost* is not instantiated in this or any other nomologically possible world (but

---

<sup>1</sup> That is, let's treat 'unicorn' as a natural kind term—a nonexistent but nomologically possible species of animal with the familiar appearance. We ignore the mythical, nomologically impossible properties. (But see Kripke 1972/1980: 156 f., who argues that—even if unicorns are nomologically possible—there are no possible circumstances in which it is correct to say that unicorns exist.)

<sup>2</sup> Segal (2000: §2.2) doesn't share my intuition that ghosts are nomologically impossible. He considers that unicorns and ghosts are both nomologically possible.

presumably is instantiated in more distant worlds with different physical laws from our own).

#### ROUND SQUARE

Round squares are logically impossible, so there is by assumption no such property as *round square* in any possible world.

How can Fodor's account deal with such cases?<sup>3</sup> The first option available to him would be denial. That is, he could claim that in some or all of these cases, we do not in fact possess the concept in question. This would not be a very successful strategy, however: clearly, you can ask whether echidnas are fierce, predict that dodos could be recreated from their DNA, wonder if schmoos would have made good pets, long to ride a unicorn, reassure yourself that ghosts don't exist and know that round squares are logically impossible. In each case, being able to do so requires being able to think about the entity in question, which requires possessing a mental representation (that is, a concept) in each case.

So it seems we must grant possession of some concept in each of the above cases. The challenge for informational semantics is to explain in each case how we can be locked to the property in question (if there is one). This involves two separate questions, the metaphysical question of whether there can be a locking relation with the property in question, and the psychological question of how we get ourselves into such a relation. To investigate this, we first need a more detailed understanding of the ontology, which will enable us to answer the metaphysical question. We will then look at the psychological question.

---

<sup>3</sup> Scott (2002, 2003) discusses a similar range of non-referring concepts and related cases of what he calls 'unacquainted content'. However, he fails to make the critical distinction between what's *uninstantiated* and what's *nomologically impossible*. The conclusions he draws are therefore rather different from mine: he claims that Fodor has to abandon atomism for any non-referring concept (and, what's more, for any concept that has been acquired other than by exposure to instances of the entities that fall under it); this is clearly not the case. (Cf. footnote 11 on p. 132 below.)

#### 4.1.2. *Ontological issues*

Assuming that at least some properties exist,<sup>4</sup> the most basic theoretical distinction in ontology is between “property minimalism” and “property realism” (although there is also a fertile middle ground). Property minimalists (such as Armstrong 1997) require that properties must be instantiated. Property realists (such as Jubien 1989), by contrast, allow that properties have an existence independently of their instances, and hence that there can be uninstantiated properties—or even that every possible property exists.<sup>5</sup>

While property minimalism is admirably austere, it won't do the work required to support an informational semantics. Since on an informational-semantic view primitive concepts get their content from nomological links to the properties they express, property minimalism would imply that we cannot have primitive concepts for (de facto) nonexistent entities. This is not plausible. The strongest version of the minimalist thesis holds that properties only exist at the moments when they are instantiated. This would have all sorts of highly implausible consequences for informational atomism, for example that people ceased to have the concept DODO at the very instant that the last dodo expired. Even on weaker versions of the minimalist thesis such as that put forward by Armstrong (1997), according to which a property exists if it has any instantiation in space-time,<sup>6</sup> there are implausible consequences for informational semantics. For such an account would imply that we could not have the unstructured concepts UNICORN, DRAGON, or any other primitive concept for a possible but non-actual entity.

We are now in a position to consider the difficulties raised by the range of concepts discussed in the previous section. Fodor, of course, is a property realist. As

---

<sup>4</sup> Which not everyone does, but never mind. (See Quine 1948/1953 for discussion.)

<sup>5</sup> For discussion of the various contemporary theories of properties, see Swoyer (1996, 2000).

<sup>6</sup> Since properties can themselves instantiate other properties, weak minimalism in fact needs to allow that a property exists if it is part of a (possibly singleton) chain of instantiations that terminates in an instantiation in space-time. See Swoyer (1996: 243).

he is careful to point out,<sup>7</sup> being locked to a property requires that the property exists, but this does not require that it is instantiated. So there is no (metaphysical) reason why we cannot be locked to *being an echidna*, *being a dodo*, *being a schmoo*, or even *being a unicorn*, on the assumption that unicorns are nomologically possible. In all these cases, there is a property to get locked to, even if it is not instantiated. These cases therefore raise no special metaphysical problem for informational semantics.<sup>8</sup>

It is metaphysically impossible, by contrast, to be locked to the property of *being a round square*, since (even for property realists) no such property exists. But recall that informational semantics is only a theory of the content of *primitive* concepts. Complex concepts get their content by composition from the contents of their constituent concepts, not from their nomological relations with properties. So, as Fodor (1998a: Appendix 7A; 2000a: 364) points out, it is perfectly possible to have the concept ROUND SQUARE, so long as it is complex (that is, made up of ROUND plus SQUARE). All that informational semantics rules out is the primitive concept ROUNDSQUARE (or any other primitive concept for a logically impossible property).

The only case we have not covered is the concept GHOST. The property of *being a ghost* is not logically impossible, and so presumably there could be such a property. In this case, however, it is no accident that the property is uninstantiated in our world, but a point of nomological necessity. This creates a problem. For if ghosts are nomologically impossible, then it's difficult to imagine how there could be (nomological) laws involving the property of *being a ghost*. To see this more clearly, consider how such laws are cashed out. One way to cash out such a law is in terms of counterfactuals. That is, we would say something like: ghosts would cause tokenings of GHOST if there were any ghosts (in the same way that we might say: dodos would cause tokenings of DODO if there were any dodos). But there can't be any ghosts in

---

<sup>7</sup> See, for example, Fodor (1987: 163, fn. 5; 1990b: 100 f., 116, 123 f.; 1994: Appendix B; 1998a: Appendix 7A; 2000a: 364).

<sup>8</sup> Viger (2001) misses this point in his thought experiment about a virus which wipes out all dogs and hence (so he claims) breaks the nomological relation between *doghood* and DOG.

point of nomological necessity, so (unlike with dodos) it doesn't make much sense to talk about what ghosts would cause if there were any. Another way to cash out the relevant law is in terms of possible worlds. On such a formulation we would say something like: ghosts cause tokenings of GHOST in worlds near us where there are ghosts, and there is no nearer world where something else does.<sup>9</sup> The problem is that the only worlds where there are ghosts are worlds where the laws of nature are different, and therefore whatever laws we formulate for this world cannot be taken to apply to them. That is, by assumption the laws of nature are conserved only in nomologically possible worlds, but ghosts don't exist in any nomologically possible world.<sup>10</sup> So it doesn't seem there can be laws about ghosts. In which case, informational semantics says that we cannot have a primitive GHOST concept. This means that GHOST must be a complex concept, like ROUND SQUARE.<sup>11</sup> This is counterintuitive, however—GHOST certainly doesn't *seem* complex in the way that ROUND SQUARE is. More importantly, it raises the question of what the primitive, nomologically possible constituents of GHOST are. In the case of ROUND SQUARE, this is obvious, but in the case of GHOST it is not. In fact, it looks like Fodor is going to have to posit some sort of *definition* (or pseudo-Russellian description) for 'ghost'

---

<sup>9</sup> Cf. Fodor (1990b: 95, 100 f.), but note that Fodor is dubious of possible-worlds analyses. See also Cain (2002: 122).

<sup>10</sup> Some philosophers believe that the laws of nature are metaphysically necessary (see Drewery 2005 for a critical discussion of this view). In this case, the only metaphysically possible worlds are nomologically possible worlds, and therefore there would be no such property as *being a ghost*. Again, the implication would be that GHOST could not be a primitive concept.

<sup>11</sup> It is important to note the distinction between the claim being discussed in the text that concepts for *nomologically impossible* entities must be complex, and the broader claim that concepts for *uninstantiated* entities must be complex. Such confusion could arise (cf. footnote 3 on p. 129 above), because Fodor was at one point tempted by (but stopped short of endorsing) the idea of a "mixed informational theory", which adds to the asymmetric dependence theory the condition that some concept tokens must actually be caused by instantiations of the corresponding property (Fodor 1990b: 119–124; see also Dretske 1981: 222 f., who proposes a similar idea). This has the consequence that concepts for uninstantiated entities must be complex (for discussion, see Baker 1991, Boghossian 1991, Cain 2002: 141 and Scott 2003: 40 ff.). He has not revived this idea in later work, although he does discuss the empiricist version of this proposal—that concepts which can't be copied from experience must be complex—in Fodor (2003a: 29, 63, 116–118).

(cf. Cain 2002: 141; Scott 2003: 40 ff.). Thus, he could propose that GHOST is actually a complex concept like INCORPOREAL PERSON, but this raises a whole host of problems that definition theory led to, and which informational semantics was supposed to avoid. And notice that Fodor will have to take the same approach for any concept of a nomologically impossible entity, of which there are plenty (think of your favourite fairy story or science fiction novel; see also Segal 2000: Chapter 2, Scott 2003). Are all of these concepts really complex? There is also a more general issue here, which is that Fodor's strategy is to cede the problematic cases to non-atomism—something that results in his theory losing its generality and appeal.

Notice that the above discussion has only dealt with the metaphysical question of whether there can be a locking relation in the cases under consideration. But the point of the SCHMOO example with which we opened this chapter was not metaphysical, it was psychological. That is, it did not question the metaphysical possibility of being locked to the property *schmoo*, but rather questioned how we could come to be in such a locking relation to a property that neither we nor anyone else had encountered. In what follows, I will show that psychological considerations suggest a solution to this problem, and that this solution also suggests a different account for concepts such as GHOST which can avoid the problems raised above. First, though, a short digression.

While we are discussing ontological matters, it is worth briefly considering the question of identity conditions on properties. One popular view is that properties are identical if and only if they are necessarily coextensive.<sup>12</sup> On this view, the property *triangular* and the property *trilateral* are the same property. This would rule out the possibility that we could have distinct atomic concepts TRIANGLE and TRILATERAL,

---

<sup>12</sup> Such a view is endorsed, for example, by Jackson (1998: 125–127) and by Fodor himself in his more recent externalist phase (1994: 61; but cf. 1987: Chapter 2; 1991). Note that we must distinguish the present view that necessarily coextensive properties are identical from the view that necessarily *coinstantiated* properties are identical. Fodor (1994), for example, accepts the former view, but not necessarily the latter (see the discussion of Quine's inscrutability problem in §2.5.2 above).



or any other distinct atomic concepts that were necessarily coextensive.<sup>13</sup> This is implausible given the powerful arguments in Sober (1982) that certain necessarily coextensive properties can confer *different* causal powers on their instances, and hence should be regarded as distinct properties.<sup>14</sup> Sober and others suggest that since the notions of coextensivity and causal powers pull apart, properties should be taken to be identical if and only if they confer the same causal or nomological powers on their instances (see Achinstein 1974, Sober 1982, Vallentyne 1998).<sup>15</sup> If Sober's arguments are correct, this would enable us to distinguish necessarily coextensive properties and hence allow distinct atomic concepts such as TRIANGLE and TRILATERAL (although this would still not allow us to account for the fact that these concepts may have different possession conditions—see §2.5.2 above).

#### 4.1.3. *Psychological issues*

The above discussion of the metaphysical issues relating to locking has shown that there is no metaphysical reason why we cannot possess any of the concepts that we considered at the beginning of this chapter. In two cases, however, informational semantics requires that these concepts be complex (that is, concepts such as GHOST that express nomologically impossible properties, and concepts such as ROUND SQUARE that express logically/metaphysically impossible properties). Next, we need to address the psychological question of how, in point of fact, someone could come to be locked to the properties in question in the various cases considered above. This

---

<sup>13</sup> If we allow a degree of atomism by postulating complex modes of presentation in these cases (while maintaining atomic *content*), this allows TRIANGLE and TRILATERAL to be distinct concepts (see §2.5.2 above; see also §4.2 below, where possible ways around this are discussed). Note that the present view allows there to be distinct atomic concepts for *contingently* coinstantiated properties such as *renate/cordate* (terms introduced by Quine to mean “creature with kidneys” and “creature with a heart”, respectively), since these are nevertheless different properties.

<sup>14</sup> Clearly, this is not to say that *all* necessarily coextensive properties should be regarded as distinct: plausibly, *being water* and *being H<sub>2</sub>O* are the same property.

<sup>15</sup> Fodor (1987: Chapter 2; 1991) also subscribed to such a view, back when he was an internalist. See also footnote 12 (p. 133) above. Note that there are important differences between the formulation of identity in terms of causal powers and the formulation in terms of nomological powers (for discussion of this distinction, and arguments in favour of the latter formulation, see Vallentyne 1998).

was the question raised by the SHMOO example in Keil & Wilson (2000: 316), and a similar problem arises for the other primitive concepts we looked at.

First, note that if GHOST and ROUND SQUARE are complex, then there should be no difficulty in explaining how we could get locked to the properties corresponding to their primitive constituents, since these are presumably all regular, locally instantiated properties (*incorporeal* and *person*, say, and *round* and *square*).

In the case of UNICORN, as we have seen above, there is a property to get locked to (on the assumption that 'unicorn' is a nomologically possible natural kind term). The only question is how we could do so, given that the property is not instantiated in our world. We are assuming, however, that although the individual cannot have had any experiences with instances of *unicorn*, they have had experiences with representations of unicorns (pictures or detailed descriptions in story books, say). It is therefore plausible that the individual has acquired a detector for unicorns, and that it is in virtue of this detector that they are locked to the property *unicorn*. This situation is no different to that for any other natural kind term: we acquire detectors for many kinds before having any actual experience of those kinds, and we are easily able to recognize instances when we come across them.

In the other cases (that is, ECHIDNA, DODO, and SCHMOO), we had assumed that the individual had had no experience of the entities in question, hence no exposure to instantiations of the corresponding properties. We had further assumed that they had had no experiences of representations of these entities. In such circumstances, the question arises how the individual could come to be locked to the properties in question. Without an explanation for this, informational semantics will be unable to explain how the individual possesses the concepts in question.

One possible explanation, of course, would be that we had innate detectors for the entities in question. It may indeed be plausible to postulate such innate detectors in some cases. For example, humans and other animals apparently have innate detectors for snakes, allowing them to be locked to the property *snake* and therefore in possession of the concept SNAKE without needing to have had prior experiences of

snakes (Tooby et al. 2005). Similar innate mechanisms probably exist to detect a range of distinctions or categories. For example, in humans, innate capacities to distinguish sex, age, race, contaminated food as well as detectors for various more specific dangers (snakes, spiders, vertical drops) have been proposed.<sup>16</sup> These may well be rudimentary detectors which are refined through subsequent exposure to instances. They are the result of natural selection and they detect properties which were important for survival in the environment in which humans evolved (the so-called ‘environment of evolutionary adaptedness’). The cognitive inflexibility that results from having such detectors hardwired (for example, in the case of the snake detector, frequent ‘false positive’ responses to curved sticks, or lack of utility in environments where there are no dangerous snakes) must be outweighed by the benefits of not having to acquire the detector in question (because the bite of a poisonous snake would frequently kill an organism before it acquired the necessary detector, say). In general, innate detectors will thus be for adaptively significant and relatively stable features of the ancestral environment.

It should be fairly clear that innate detectors for echidnas, dodos and schmoos are implausible. Apart from being of inherently low significance for the survival of ancestral humans, none of these species would have been stable features of the environment of evolutionary adaptedness in any case—echidnas and dodos were extremely geographically restricted and schmoos were by hypothesis nonexistent. There is therefore no reason for the relevant innate detectors to have evolved. In fact, it would only be reasonable to posit innate detectors in these cases if we assumed that *virtually all* concepts were innate—the kind of radical concept nativism that Fodor has proposed in the past, and which his more recent thinking is an explicit attempt to move away from.

---

<sup>16</sup> See Tooby et al. (2005) and Sperber (2005), as well as the references cited therein. While most of these capacities plausibly evolved to perform the function that they do, this is unlikely in the case of race, which is more likely to be a by-product some other evolved capacity, such as coalitional psychology (see Cosmides et al. 2003). Most of these capacities are visual, but there is also evidence of innate auditory and pheromone detectors in a range of vertebrates (Sewards & Sewards 2002).

So it seems that our hypothetical individual cannot have innate detectors for echidnas, dodos and schmoos, and nor could they have acquired such detectors, given that they have had no experiences of the entities in question and no experiences of representations of these entities. But the individual can think about these entities, and so must be credited with the corresponding concepts. It follows that the individual must be locked to the relevant properties not via a perceptual detector, but in some other way.

This in itself is not a surprising conclusion. After all, there are various kinds of properties that we do not get locked to via direct perception. Very often we may have to consciously employ encyclopaedic knowledge, or an actual encyclopaedia, to identify something. Think of bird watchers: those species that they have perceptual detectors for can be identified automatically without the need for conscious thought, whereas other species can only be identified by consulting background knowledge or a bird book. Or consider Putnam's famous example about elms and beeches (Putnam 1975). Most of us cannot tell the difference between these two species of deciduous tree. We have the concepts ELM and BEECH, but no separate detectors, just a more general detector (for large deciduous trees, say). There exist tree experts, however, whose perceptual detectors are more finely developed, and who can tell the difference just by looking. In other cases, even experts do not have the relevant perceptual detectors. For example, humans do not have the perceptual capacities to directly observe protons or x-rays, but whether we are laypeople or scientists we can have the concepts PROTON and X-RAY.

Fodor allows that there are various different ways in which we can get locked to particular properties (see Fodor 1998a: 75–80; see also Margolis 1998). That is, there are various different mechanisms which can ensure that Xs reliably cause us to token the concept X. Having a detector for Xs is one obvious way, but there are others. For example, scientists develop detailed theories which predict how entities that cannot be directly perceived behave and how to detect them. The fact that a particular pattern in a bubble chamber causes a scientist to token PROTON can then be explained

by the fact that the scientist has a theory about protons which (together with some experimental apparatus) mediates between protons and tokenings of the concept PROTON, and hence allows the subject to be locked to the property of *being a proton*. Another mechanism which could sustain the locking relation is a disposition to defer to experts, as suggested (in a somewhat different context) by Putnam (1975) and Burge (1979). If we are prepared to defer to an expert about which trees are elms and which are beeches, or about the presence of protons and x-rays, then the fact that these entities reliably cause tokenings of the corresponding concept in experts will ensure that they reliably cause tokenings of the corresponding concept in us as well.

This is all very plausible, and I think that it allows Fodor to account for the facts in the ‘schmoo’ example raised by Keil and Wilson, as well as the ‘echidna’ and ‘dodo’ examples we have been considering. There is something unsatisfying about it, though. We seem to be rather spoiled for choice when it comes to locking. There are *so many* ways that we can get locked to a property, including just having a generalized disposition to defer to others who may be more knowledgeable (we don’t even need to know what kind of expert we should defer to in a particular case: we can use the Yellow Pages or the human-expert equivalent, who will in turn defer to another expert). Such an account can do some heavy-duty metaphysical work, but it leaves open the question of how we actually think with the concepts in question (which is ultimately what concepts are *for*). Deference to experts may be fine as a metaphysical strategy, but is it of extremely limited utility for narrow psychological purposes.

The approach that I have been developing can shed some light on this. Intuitive inferences rely on meaning postulates attached to concepts. But deference can’t explain how we come to have meaning postulates, and therefore deference doesn’t allow a concept to be employed in intuitive inference. This shows the psychological limitations of deference. But it raises a metaphysical problem at the same time.

Content, I have suggested, is partly constituted by nomological mind–world links in the way proposed by Fodorian informational semantics. However, *contra* Fodor,

content is also partly constituted by meaning postulates, in the sense of mental rules of inference. It is these meaning postulates that account for the content of logical terms, and, more generally, the logical/inferential properties of all concepts. Our mental ‘deductive device’ also employs these meaning postulates in performing intuitive inference. The problem, then, is clear: to account for concept possession, we need not only an account of a plausible sustaining mechanism, but also an account of how we come to have the meaning postulates that are partly constitutive of the content of the concept in question. Whatever form this account takes (something we will come back to in chapter 5), it is *prima facie* implausible that deference, for example, will have any role to play. After all, it makes sense to defer to experts in *detecting* Xs, but we can’t very well defer to experts when we make intuitive (and generally subconscious) *inferences* about Xs (even if we’re happy to let experts—and their graduate students—do all the reflective, scientific thinking about Xs). Sociology is one thing, cognitive psychology quite another.

Well, but if conceptual content, at least for many concepts, consists of two separate aspects, and if the mechanisms that mediate one aspect don’t necessarily mediate the second aspect, then it is plausible that in many situations we could have one part of the content without having the other part—that is, we could be locked to *Xness* but not have the canonical X-involving meaning postulates, and therefore be in possession of a defective X-concept. We might, for example, be able to (eventually) find out when there were Xs in the vicinity, but not be able to think intuitively about Xs.<sup>17</sup>

By drawing on the distinction between intuitive and reflective concepts that has been developed by Dan Sperber and others (for example, Sperber 1994, 1996, 1997; Cosmides & Tooby 2000a), we can develop a rigorous account of how such

---

<sup>17</sup> The converse is also possible: we could have meaning postulates for a concept X without being locked to *Xness*. Suppose we are told that agoutis eat brazil nuts, but have no idea what an agouti is. We could infer that an agouti is a kind of animal, and thus start to construct a schematic AGOUTI concept, including the meaning postulate ‘ $\phi$  AGOUTI  $\psi \rightarrow \phi$  ANIMAL  $\psi$ ’, but having no referential content. See the discussion of reflective concepts in §§4.3.2 and 4.5 below, as well as chapter 5.

defective concepts could arise and the role they play in our psychology. Before switching to a psychological mode, however, I want to briefly discuss the question raised earlier (§4.1.2, fn. 13) and in chapter 3 (§3.3) of how coextensive concepts can be type-distinct.

#### **4.2. Mentalese orthography**

Recall that, according to Fodor, there are two aspects to concept-individuation: the content of the concept, and its mode of presentation. Contents are individuated with respect to the properties they are nomologically linked to. Modes of presentation are individuated with respect to their causal roles. The computations of the representational mind are blind to both of these methods of individuation, however. If we accept Turing's fundamental insight, then processes of mental computation must be *syntactic*, and therefore cannot be sensitive to mind-world relations as such or to causal roles as such. This implies that if we want our mental processes to respect the relevant semantic/causal distinctions (for example, if we want them to be truth-preserving) then there had better be some *formal* properties of mental-representation types which can ensure that syntax respects semantics (Fodor is forever stressing this point; see, for example Fodor 1982; 1987: 18 f.; 1998a: Chapter 1).

What could these formal properties be? It is obvious in the case of complex expressions of Mentalese that their compositional constituent structure must be such that the correspondence of syntax and semantics is ensured (see the various papers in Fodor & Lepore 2002). But what about primitive Mentalese expressions? There must be some causally-relevant formal property that distinguishes them—their “shape”, as one says—since they don't have distinct syntactic properties. Unlike natural language, we obviously cannot fall back on phonology, and it does not seem that Mentalese can have an orthography in the sense of natural language. To speak of a Mentalese orthography would imply that there must be a “reader” to interpret this orthography (cf. Millikan 1993; Fodor 1994: Appendix A). This is not to say,

however, that mental representations have no relevant formal properties. The issue is what the formal properties of token mental representations could be such that they are treated by the mind's computational machinery as being of a particular type.<sup>18</sup> Candidates for the relevant formal properties could include neurological properties (Fodor 1994: 107), some other higher-order physical property (Fodor 1987: 18, 156 fn. 5), or perhaps some more abstract cognitive property such as an 'address' to a location in memory (Sperber & Wilson 1995: 86). Primitive Mentalese expressions will be distinguished on the basis of their formal properties. Thus, the Mentalese expression CAT will be formally distinct from the Mentalese expression BANANA because the two expressions will have different "shapes", or will address different locations in memory. But what about primitive expressions of Mentalese that are coreferential? It is clear that we also need some way to distinguish these: notoriously, CICERO and TULLY can be different concepts, but would plausibly correspond to Mentalese primitives (see §3.3 and Fodor 1994: 106). We therefore need some formal difference that the mind can be sensitive to in order that CICERO and TULLY can count as different expressions of Mentalese, while at the same time allowing that someone who knows that they refer to the same individual will treat them identically. For example, we could allow that these two expressions have different "shapes", but that these "shapes" would participate in identical causal interactions in the mind of a person who treated them coreferentially; similarly, we could allow that they address different locations in memory, but that these locations could be cross-referenced. This is complicated, of course. Plausibly, every mental representation token will differ from every other along some physical dimension.

---

<sup>18</sup> As always, it is important to distinguish psychology from metaphysics. The "shape" of a mental representation is responsible for how that representation is treated by the mind's computational machinery. But the metaphysics of representation typing need not (and almost certainly does not) depend on the same property of "shape". It is generally accepted that type physicalism is false, and if this is the case—that is, if the same representation type can be tokened in different physical media—then we need something other than physical "shape" as a metaphysical basis for representation typing. Perhaps functional role can do this (or perhaps not). See Fodor (1994: Appendix A) and Aydede (2000).



What we need is a physical property that performs the equivalent of an “orthographic” function. Memory addresses do better since there will be some reliable physical basis underlying the mind’s organization of its information, but we still need a rigorous account of this, including as regards cross-referencing. For discussion of some difficulties raised by such an approach, see Millikan (1993). Much more needs to be said, but I am less pessimistic than Millikan about the prospects.

### **4.3. Ways of believing**

#### ***4.3.1. Intuitive and reflective beliefs***

Human beings have beliefs about a huge range of things. We believe that chickens are animals; that unsupported objects fall downwards; that people’s actions are generally motivated by their beliefs and desires; that tables only move if an external force is applied, but that the same is not true of chickens. We may not necessarily be consciously aware of such beliefs, or of their justification, but we hold them nonetheless (they have been shown to govern many of our everyday interactions with people, chickens and tables), and by-and-large they tend to be veridical and mutually consistent. We may also believe that water consists of hydrogen and oxygen; that astrologers can predict the future; that clocks in orbit run more slowly than those on the ground; that the common cold results from exposure to cold temperatures; and that the common cold is the result of a viral infection. We are generally aware that we hold such beliefs (we often invoke them in explanations of our behaviour), as well as the basis or authority on which we hold them, but they do not show a high degree of veridicality (such as regarding the abilities of astrologers) or mutual consistency (such as regarding the causes of the common cold).

As Sperber (1975, 1996) notes, when faced with such facts, as theorists we have a number of choices. We could abandon the idea that people tend to be (if not perfectly, then at least reasonably) rational. Few people, and fewer psychologists,

find such an approach appealing. Alternatively, we could abandon objectivity, and claim that different cultural ‘realities’ give rise to different criteria for rationality. This has been a more popular approach, but raises major problems of its own, particularly for those with naturalistic tendencies: naturalism demands an objective reality to be naturalistic about. Sperber endorses a third view. He introduces a distinction between two kinds of belief that we can have—‘intuitive’ beliefs and ‘reflective’ beliefs—and argues that these beliefs achieve rationality in different ways. This can explain the relevant facts, without the need to deny human rationality or succumb to cultural relativism.<sup>19</sup> I will briefly summarise Sperber’s arguments for positing two distinct kinds of belief, and then look at how this might be relevant to the discussion in §4.1.

In order to account for belief/desire psychology we might (following Schiffer 1981 and much subsequent literature) postulate two mental boxes for the storage of representations: a ‘belief box’ and a ‘desire box’.<sup>20</sup> Representations stored in the belief box are treated by the mind as representations of actual states of affairs, and those stored in the desire box are treated as representations of desirable states of affairs. We could also postulate other ‘boxes’ to account for other propositional attitudes (hope, fear, doubt, regret and so on), but given the indefinite variety of

---

<sup>19</sup> There have been some related proposals put forward in the literature. Fodor (1984c) makes a similar suggestion that there are two distinct routes to belief fixation—observation and inference—which implies a corresponding taxonomy of beliefs. Fodor stops short of fully endorsing this view, however. We will look at how this relates to Sperber’s distinction in §4.3.2 below. Also, a number of theorists (Epstein et al. 1992, Epstein 1994, Sloman 1996, Smith & DeCoster 2000, Lieberman et al. 2002, Schneider & Chein 2003) have put forward ‘dual-process’ theories of cognition, which bear a superficial similarity to Sperber’s proposal (noted by Pyysiäinen 2003), but which differ in fundamental respects.

<sup>20</sup> Attitude boxes are of course just a convenient metaphor (and certainly not a neurophysiological proposal). The idea behind this metaphor is that attitudes are typed with respect to their functional properties (for example, desires tend to cause action designed to bring about the conditions for their satisfaction). So beliefs and desires are just tokens of propositions entertained in such a way as to bestow on them the right causal powers: being in a particular location, having the right tag attached, or whatever the cognitive architecture may require (Cosmides & Tooby 2000a, for example, propose that belief is the unmarked architectural default, so that representations that are not explicitly marked in some other way are treated as beliefs).

possible attitudes, it may be that in many cases these are metarepresentational rather than architectural (see Sperber 1997; Sperber & Wilson 1995: 73–75). According to this proposal, doubting that agoutis like milk could come about through having in one’s belief box a representation such as ‘I DOUBT [THAT AGOUTIS LIKE MILK]’.<sup>21</sup> This is a metarepresentation because the representation expressing the content of the attitude (that is, AGOUTIS LIKE MILK) is embedded within the attitude representation (that is, ‘I DOUBT THAT *P*’). It seems undeniable that humans have such metarepresentational capacities—for example, we clearly have the ability to consciously reflect on the contents of our own beliefs and desires, as well as manipulate metarepresentations in more automatic and subconscious ways (such as in understanding utterances or in mind-reading).<sup>22</sup> It follows that it is possible for us to entertain in this metarepresentational way any attitude that we have the conceptual vocabulary to represent. A more limited range of attitudes (perhaps just belief and desire) can arise from the basic cognitive architecture, rather than metarepresentationally.<sup>23</sup>

This raises the possibility that the same attitude can be entertained in some instances architecturally, and in other instances metarepresentationally. For example, metarepresentation allows us to entertain beliefs without the embedded representation itself appearing in the belief box. Furthermore, metarepresentation also allows us to entertain a whole range of attitudes which, although not strictly attitudes of belief, would tend to support the truth of their propositional objects, as in (1)–(4):

(1) EVERYBODY KNOWS [THAT REAL TRUFFLES ARE EXPENSIVE]

---

<sup>21</sup> There are other possibilities: having the belief PROBABLY NOT *P*, for example, would also give rise to the doubt that *P*.

<sup>22</sup> For detailed discussion, see the papers in Sperber (2000).

<sup>23</sup> Note that we cannot do without any boxes at all, by treating all attitudes metarepresentationally. If there was no ‘architectural’ belief, then any metarepresentation could be entertained in any number of ways (“I BELIEVE THAT *P*” can be doubted, desired, feared and so on, just as *P* can be), and further metarepresentational embeddings merely create an infinite regress of this problem. Whether we have any other boxes in addition to the belief box is an empirical question, however. See Sperber (1997: 68).

- (2) IT IS A SCIENTIFIC FACT [THAT A GLASS OF WINE A DAY IS GOOD FOR THE HEART]
- (3) THE TEACHER SAID [THAT THERE ARE MILLIONS OF SUNS IN THE UNIVERSE]
- (4) IT HAS BEEN PROVEN [THAT COMMUNISM DOES NOT WORK]

(adapted from Sperber 1997: 70–76)

Sperber refers to these as ‘credal attitudes’. He generalises the situation as follows. A belief with metarepresentational content may provide a validating context (V) for the embedded representation (R). In such a situation the individual has two credal attitudes, one with content V(R) (in virtue of the representation V(R) occurring in their belief box), and one with content R (not in virtue of the representation R itself occurring in their belief box, but in virtue of R occurring in a validating context). There is no fixed or well-defined range of validating contexts, which can include, for example, reference to authority, reference to scientific theory, explicit proof, or divine revelation. Sperber refers to the attitude of belief arising from a representation occurring in the belief box as ‘intuitive belief’ and the attitude of belief arising from a representation occurring in a validating context as ‘reflective belief’. There is a certain sense in which reflective beliefs encode aspects of their etiology in the form of their validating context.<sup>24</sup>

There is one more refinement that we need in order to complete this picture. Consider the belief that there are no elephants in Antarctica. Assuming that we have never previously considered the matter of elephants in Antarctica, we would not want to propose that we have a token of the relevant proposition in our belief box. But we would certainly assent to the proposition if it was presented to us. There would in fact seem to be a vast number of beliefs of this kind that are merely

---

<sup>24</sup> There are clear resonances here with the distinction between episodic memory (memories of past events or episodes) and semantic memory (knowing that). See Tulving (1972); see also Smith & DeCoster (2000). It need not always be the case, however, that a validating context encodes etiology: the attitude of belief itself could be entertained metarepresentationally, by having a representation of the form ‘I BELIEVE THAT P’ in one’s belief box. This might be the case for certain delusional beliefs (for example, the belief of a person with the Cotard delusion that they are dead, which it is difficult to see as anything other than a reflective belief). Cf. Currie (2000: §4).

dispositional, rather than actually represented (see Dennett 1975; Fodor 1984b/1990: 321; Sperber 1996: 86 f.). Since dispositional beliefs are not explicitly represented, they cannot be said to be either intuitive or reflective; they are only potential beliefs. But if they can be derived from our existing intuitive beliefs by spontaneous inference,<sup>25</sup> this would explain the disposition (that is, it would explain how we could come to represent them as intuitive beliefs).

This approach can go some way to accounting for the rationality of intuitive beliefs (see Sperber 1996: 87). Suppose that the basic mechanism by which a representation gets added to the contents of the belief box is via perception.<sup>26</sup> All other beliefs are directly or indirectly derived from these perceptual beliefs by the operations of spontaneous inference mechanisms (our deductive device). The fact that perception is (thanks to evolution) generally veridical makes this a reasonable strategy. And the fact that the rules of inference on which the deductive device operates (that is, meaning postulates—possibly supplemented by some non-demonstrative inferential procedures) are generally valid, hence truth-preserving, ensures that adding derived propositions directly to the belief box is also reasonable. The only reason not to do so would be if a derivation resulted in a conclusion which contradicted an assumption being held in the working memory of the deductive device.<sup>27</sup> In such a situation, one of a variety of (possibly conscious) procedures is initiated to resolve the contradiction (for example, by rejecting the proposition which is ‘weaker’ in some well-defined sense, or by searching encyclopaedic memory for evidence for and against the contradictory propositions; see Sperber & Wilson 1995: 95, 114 f. for detailed proposals). Such a contradiction-resolving procedure will tend

---

<sup>25</sup> See §2.2 above; Dennett (1975); Sperber & Wilson (1995: Chapter 2).

<sup>26</sup> “Seeing is believing”, after all; but on this account, so is hearing, feeling, smelling, tasting (and possibly introspecting—Sperber 1996: 87 proposes that our reflective awareness of our own mental states counts as perception for present purposes).

<sup>27</sup> Of course, it would be even better for the deductive device to check the consistency of every derived proposition with all other propositions in the belief box. Practical limitations would rule out the possibility of such a mechanism, however.

to increase the mutual consistency of intuitive beliefs, and hence provide a further explanation for the rationality of intuitive belief.

We now have the following characterization of intuitive and reflective beliefs:

(5) **Intuitive beliefs**

Beliefs derived from perception (including introspection), as well as beliefs derived directly or indirectly from these by the operations of the deductive device (on the basis of meaning postulates/procedures attached to concepts). Introspection may deliver metarepresentational beliefs of the form ‘I BELIEVE THAT *P*’, or of the form *V(R)* more generally, in which case only the meta-belief is intuitive.<sup>28</sup>

**Reflective beliefs**

Beliefs entertained not in virtue of occurring in the belief box, but in virtue of being embedded in a validating context. Such beliefs may be derived from communication (JOHN [WHO I TRUST] SAID *P*) or conscious thought (THERE IS EVIDENCE IN SUPPORT OF *P*).<sup>29</sup>

Intuitive and reflective beliefs achieve rationality in different ways. Intuitive beliefs do so by being derived from perception and intuitive inference and by being mutually consistent. This explains why many beliefs (about the behaviour of animals, people, and inanimate middle-sized objects, say) show very little cross-cultural variability.<sup>30</sup> Reflective beliefs achieve rationality on the basis of their validating context. They are therefore likely to vary widely depending on socially-

---

<sup>28</sup> Unless, of course, the embedded proposition is represented independently in the belief box. The point is that embedded propositions are thereby ‘insulated’ from intuitive thought processes, and are not automatically disembedded and themselves added to the belief box (see below).

<sup>29</sup> In addition to beliefs, a whole range of other reflective attitudes can be entertained in virtue of a metarepresentation of the appropriate form—such as ‘I DOUBT/FEAR/REGRET THAT *P*’—occurring in the belief box.

<sup>30</sup> Cf. Fodor: “...our agreement on the general character of the perceptual world might transcend the particularities of our training and go as deep as our common humanity.” (1984c: 40).

religiously- and culturally-determined views concerning what counts as a validating context (the gods, sacred texts, gurus, science, and so on). This explains the huge diversity of apparently contradictory beliefs observed across cultures.

A question arises as to how reflective beliefs can become intuitive. One hypothesis would be that all reflective beliefs, R, are automatically disembedded, or 'disquoted' to use Sperber's term,<sup>31</sup> from their validating context V(R) and added directly to the belief box. If this is the case, it would clearly undermine the basis for drawing a distinction between intuitive and reflective beliefs. Sperber (1996, 1997) provides two arguments against this hypothesis. First, he notes that such a strategy would be poor cognitive design. We often have credal attitudes to propositions that we do not fully understand (such as scientific or religious doctrines). Such information can be useful, but it would be dangerous to integrate it into our intuitive knowledge-base in an unrestricted way. A more compelling consideration, however, is that expressions can occur in reflective beliefs which could not occur in intuitive beliefs. This is most obvious in the case of language comprehension. In order to understand an utterance, we must be able to form a corresponding mental representation. This implies that we must have a way to mentally represent expressions that we do not fully comprehend. For example, we can have a thought corresponding to the sentence in (6):

(6) John said that agoutis are smaller than capybara

even if we have never come across the words 'agouti' or 'capybara' before; if we trust John's judgement in such matters, we will come to believe (reflectively) that agoutis are smaller than capybara. In order to mentally represent the propositions in question, however, we need a metarepresentational device: the Mentalese equivalent

---

<sup>31</sup> The term 'disquotation' was introduced by Quine (1974/1976) as a characterization of Tarskian T-sentences, and is widely used in philosophical discussions of the notion of truth, but Sperber uses it in a somewhat different sense.

of quotation marks (which I shall indicate with angle quotes).<sup>32</sup> We could then have the reflective beliefs in (7)–(8):

(7) JOHN SAID [THAT «AGOUTIS» ARE SMALLER THAN «CAPYBARA»]

(8) «AGOUTIS» ARE SMALLER THAN «CAPYBARA»

The conceptual equivalent of a word such as ‘agouti’ that we do not fully understand, then, can be seen as a Mentalese expression, X, marked off in some particular way (which we are indicating orthographically with quotation marks). Its mode of presentation could take one of two forms. It could be either that X is a new (that is, primitive) item of Mentalese or that it is a phrase of Mentalese made up of existing items.

Since we do not have non-metarepresentational concepts corresponding to ‘agouti’ and ‘capybara’, we cannot disquote the Mentalese expressions «AGOUTI» and «CAPYBARA» to form the intuitive belief in (9).

(9) AGOUTIS ARE SMALLER THAN CAPYBARA

Thus, (8) cannot play a role in our intuitive thinking until we have acquired the concepts AGOUTI and CAPYBARA, allowing us to disquote these concepts from their metarepresentational contexts.

We can see metarepresentation (including Mentalese quotation marks) as an important cognitive safeguard. That is, it prevents representations that we do not fully grasp, because we lack the requisite conceptual basis for doing so, from invading our intuitive knowledge-base. It also imposes an additional requirement of disquotation on beliefs acquired through communication, allowing us to consider the

---

<sup>32</sup> It may be that metarepresentation operates only at the level of whole propositions. Reflective concepts would then be those that occur in metarepresented propositions, but there would be no individually metarepresented concepts. However, I will make the assumption in what follows that it is possible to have metarepresentation at the individual concept level; nothing much will turn on this, however.



credibility of the source as well as the inherent plausibility of the reflective belief itself before we add it to our intuitive knowledge-base.

Sperber (1997: 75 f.) suggests that in addition to not having the necessary concepts, there is another potential reason why we may be unable to disquote a reflective belief. That is, although we may in some cases possess the necessary concepts (and syntax) to represent a belief, we may nevertheless be unable to grasp the content in question. The example he gives is (10):

(10) THE TEACHER SAID [THAT THE FATHER, THE SON AND THE HOLY GHOST ARE ONE]

Let us grant for the sake of argument that we do have all the necessary concepts in this case (although it is unlikely that we fully grasp religious concepts such as HOLY GHOST, a point that Sperber would certainly not disagree with), and let us assume that we trust the teacher in question as an authority on such matters. What we are being asked to consider is the possibility that, in spite of this, we may not be able to grasp the content of the belief. But if we follow this line of thinking, it suggests that there can be something *other than* its constituent concepts and their mode of combination that contributes to the meaning of an expression of Mentalese. We are in obvious danger at this point of giving up on the compositionality of thought, a very unattractive option. This is presumably not what Sperber had in mind. Rather, he seems to have been considering that although we could represent the belief, we would have no intuitive grasp of it—because it conflicts with another intuitive belief, namely that fathers must be distinct individuals from their sons (cf. Sperber 1994: 62).

Building on this idea, let us propose that in the course of disquotation, a consistency test is performed on candidate intuitive beliefs. This could be achieved if we supposed that, once disquoted, candidate intuitive beliefs are added not directly to the belief box, but instead to the working memory of the deductive device. The deductive device then follows its normal procedure, as set out in Sperber & Wilson (1995: 93 ff.) and discussed briefly above. If the device detects a contradiction, it

halts, and attempts are made to resolve the contradiction. This will result either in the candidate intuitive belief being deleted, in which case it is not added to the belief box, or in the pre-existing contradictory belief being deleted, allowing the derivation to continue and possibly (if no further contradictions arise) to the candidate intuitive belief being added to the belief box.

Note, by the way, that the existence of two kinds of belief also explains how individuals can persist in holding simultaneous contradictory beliefs (Sperber 1994; cf. Sloman 1996). Contradictions between intuitive beliefs, once these are activated simultaneously, should trigger a procedure to resolve the contradiction, as we have seen. Contradictions between reflective beliefs can be resolved through a conscious search for and evaluation of evidence in support of each. But if one of a contradictory pair of beliefs is intuitive, and the other reflective, then there are reasons to suppose that the contradiction will persist. Take, for example, the well-known Müller-Lyer illusion. On initial presentation, the lines appear to us to be of different lengths, and this is what we come to automatically and intuitively believe. If we then find out that the lines are in fact of the same length, for example by measuring them, we are able to form a reflective belief that this is the case. These two beliefs are contradictory, but neither manages to displace the other, and so the contradiction is not resolved. Looking again at the diagram will once again produce the intuitive belief that the lines are of different lengths, while at the same time we continue to consciously/reflectively believe that they are the same length.

Such a situation is to be expected on the present account: the belief produced by our perceptual mechanisms that the lines are of different lengths is an intuitive belief, and is not in contradiction with any other intuitive beliefs. Once we have measured the lines and consciously compared their lengths, we come to form the reflective belief that the lines are of the same length. If we now try to disquote this belief, however, we face a problem: the deductive device finds that it is in contradiction with the intuitive belief that the lines are of different lengths. Being grounded directly in perception, we will be extremely unlikely to reject the intuitive belief in

favour of a disquoted reflective belief.<sup>33</sup> The two belief systems therefore continue to simultaneously produce contradictory beliefs, and although we can consciously override the prepotent intuitive belief once we form the conflicting reflective belief, this does not eliminate the intuitive belief.

If, as the above considerations suggest, we can have reflective beliefs that cannot be represented intuitively, it follows that there can be no automatic process whereby reflective beliefs are disquoted and added directly to the belief box. Rather, there are at least two specific constraints on disquotation. First, we will be prevented from disquoting a belief if we do not have the primitive non-metarepresentational expressions of Mentalese required to represent it. Second, we will be prevented from disquoting a belief if it leads to a contradiction with other (/more strongly held) intuitive beliefs.

This raises the question of how we are able to make any use of reflective beliefs in reasoning. There is no doubt that it is important to be able to insulate reflective beliefs in some way to prevent them from invading our intuitive knowledge-base. But we do not want them so well-insulated that they are inaccessible to reflective thought, otherwise they would be of no use at all. Sperber (1997) suggests that reflective thought involves (at least in part) disquoting a reflective belief to a temporary buffer attached to the deductive device, what Nichols & Stich (2000) call a “possible world box”. This allows the reflective belief to be brought together with other assumptions so that inferences can be drawn, but in a way which keeps them insulated from our intuitive beliefs—the output of this process must be re-embedded in the metarepresentational context before being added to the belief box. Pretence, counterfactual reasoning and conditional planning, for example, presumably work in some such way (see Leslie 1994, Nichols & Stich 2000, Perner & Lang 2000). I take

---

<sup>33</sup> This is presumably because not only is the visual perception module in question informationally encapsulated (in the sense of Fodor 1983), but also perceptual representations (that is, the outputs of the various perception modules) are *ipso facto* accorded maximum strength by the belief fixation system. (An alternative explanation could be that the deductive device does not permit beliefs in the process of disquotation to displace beliefs that are already intuitive.)

it that there is more to reflective thought than applying our intuitive inferential mechanisms to disquoted beliefs, however. In particular, there must be inferences that we make reflectively on the basis of stored knowledge (in the form of axioms, schemas, rules of thumb, and so on) rather than intuitively on the basis of meaning postulates or procedures for spontaneous non-demonstrative inference.

#### *4.3.2. Intuitive and reflective concepts*

As is clear from the preceding discussion, in parallel with the distinction between intuitive and reflective beliefs we can draw a distinction between intuitive and reflective concepts.<sup>34</sup> Intuitive concepts are those which may appear in intuitive beliefs and take part in intuitive inference. Any of these concepts may in principle be metarepresented, as may a range of other concepts of which we do not have a full or intuitive grasp. These metarepresented concepts are the reflective concepts.

A more detailed characterization of the intuitive concepts can be given, following Sperber (1997), as follows. Recall that, for Sperber, intuitive beliefs are those beliefs that are derived from perception as well as any beliefs derived from these by our intuitive inference mechanisms. The vocabulary of intuitive beliefs is therefore the vocabulary of the conceptual representations produced as output by the perceptual modules (what I will call “perceptual concepts”), together with the vocabulary of the intuitive inference mechanisms (what I will call “inferential concepts”). This distinction between perception (/observation) and inference is a venerable—if much maligned—one in the philosophy of science (see Fodor 1983: §III.6, 1984c for discussion).

Before looking at the details of these two kinds of intuitive concept, it is worth underlining that in drawing a distinction between perceptual and inferential concepts, I am not claiming that our perceptual mechanisms do not engage in any sort of inference. One often hears it said that vision involves inference, for example when

---

<sup>34</sup> Quine (1977) draws a somewhat similar distinction between what he calls ‘intuitive’ and ‘theoretical’ concepts, and discusses the ways in which intuitive concepts develop into theoretical concepts in the course of experience, education and the development of science.

the missing parts of partially occluded objects are filled in; and it is certainly not in doubt that sensory inputs massively underdetermine conceptual outputs in the general case. While certain assumptions may be hardwired into the architecture of our perceptual mechanisms, this still plausibly leaves room for some genuine inference. This is, after all, why input modules are presumed to have proprietary *databases*; these data are the assumptions used in drawing inferences. But none of this implies that we cannot draw a different, principled distinction between perception ('what we see', a process that takes nonconceptual inputs from the senses and delivers conceptual outputs) and inference (the conclusions we draw from what we see, a process that takes conceptual inputs and delivers conceptual outputs). Fodor's modularity hypothesis supports precisely this latter distinction, as he discusses at some length; see also Sperber (1994).

#### *Perceptual concepts*

Our perceptual mechanisms can be regarded, following Fodor (1983), as a set of informationally encapsulated modules whose task it is to assign concepts to representations of sensory stimuli (the stimulus representations themselves being delivered by the sensory transducers). The idea here is not that there is one input module for each of the senses, but rather that there is one input module for each of a number of input domains, such as colour perception, shape perception, conspecific face recognition, conspecific voice recognition, and so on (see Fodor 1983: §III.1). The reason why they are *input* modules is that they are at the 'edge' of cognition, taking nonconceptual information from our sensory transducers, and assigning conceptual interpretations that can be understood by 'central' cognition. What I am calling 'perceptual concepts', then, can be seen as those concepts that are assigned to sensory stimuli by such input modules. But which concepts are these? How are we to decide which concepts are perceptual, and which are inferential?

Answering this question involves determining how "shallow" the outputs of input modules are (see Fodor 1983: §III.6). Fodor notes that if input modules are

informationally encapsulated, their outputs must be shallow enough to have been derived purely on the basis of their modality-specific sensory inputs and their proprietary computations and database. At the same time, outputs cannot be so shallow that they are still sub-conceptual<sup>35</sup> (“phenomenologically inaccessible”, as Fodor says): the whole point about perception is that it delivers conceptualized output that can be understood by central thought processes.<sup>36</sup>

Fodor’s proposal is that the output of the perceptual processes (in particular visual perception, about which the most is known) corresponds to what has been identified by psychologists as the “basic level of categorization” (Brown 1958, Rosch et al. 1976; for an overview, see Murphy 2002: Chapter 7). This is generally the most abstract level at which objects retain broad perceptual similarity—for example, shape reliably correlates with kind (Rosch et al. 1976, Landau 1994, Landau et al. 1998a). This means that objects at this level can be reliably categorized on the basis of encapsulated modality-specific processes. For example, consider a hierarchy of increasing abstractness such as *Blue Point*, *Siamese*, *cat*, *mammal*, *animal*, *physical object*. The basic level is ‘*cat*’, and this corresponds to the most abstract category in the hierarchy whose members show broad perceptual similarities (something can be cat-shaped, but not mammal-shaped). Basic level categories are acquired earlier (Horton & Markman 1980), identified faster (Intraub 1981), and tend to correspond to monomorphemic lexical items (Berlin et al. 1973).

On this view, then, our perceptual mechanisms assign basic-level concepts to sensory stimuli. The perceptual concepts can thus be broadly identified with the

---

<sup>35</sup> The three levels (primal sketch, 2½-D sketch, 3-D model) proposed by Marr (1982)—and similar interlevels of analysis in other sensory modalities—are sub-conceptual in this sense. Cf. Fodor (1983: 94).

<sup>36</sup> Notice that some such point can still be made regardless of how the debate on nonconceptual content turns out (see §3.1 above). That debate is about whether there might be nonconceptual aspects of content (perceptual experience, for example); it is common ground in the debate that some content is conceptualized (perceptual beliefs, for example).

basic-level concepts studied extensively by Rosch and others.<sup>37</sup> Sperber clearly has this sort of thing in mind: he suggests that the “mental vocabulary of intuitive beliefs is probably limited to basic concepts: that is, concepts referring to perceptually identifiable phenomena and innately pre-formed, unanalysed abstract concepts” (1996: 89). Which brings us to the other class of intuitive concept: the inferential concepts.

### *Inferential concepts*

The vocabulary of the intuitive inference mechanisms constitutes the class of inferential concepts. As the quote from Sperber above makes clear, he considers that the intuitive concepts are either perceptual, or they are innate (any concept for which we have an innate detector, such as SNAKE, would be both perceptual and innate). Thus, the vocabulary of the intuitive inference mechanisms is taken to include, in addition to perceptual concepts, innate non-perceptual concepts.<sup>38</sup> In what follows, I will use the term ‘inferential concept’ to refer just to those concepts that are introduced by inference rules (meaning postulates) and that do not occur in the vocabulary of the perceptual mechanisms. For example, suppose that LIVING KIND is a non-perceptual concept (a possibility I will discuss below), but can be introduced via the spontaneous inference ‘ANIMAL → LIVING KIND’; then LIVING KIND would be

---

<sup>37</sup> Two caveats: first, we may have perceptual detectors for more abstract categories, such as ANIMAL. But this is not inconsistent with claiming that perceptual output is at the basic level: perhaps what appears in the perceptual output in such a case is UNKNOWN ANIMAL KIND, say. Cf. footnote 49 (p. 167) below. Second, it is possible that learned perceptual expertise can extend the range of perceptual concepts beyond basic-level concepts to subordinate-level concepts. For further discussion, see chapter 5.

<sup>38</sup> It is also possible that the outputs of the perceptual mechanisms include some logico-syntactic concepts in addition to perceptual concepts. These could be introduced for example by inference rules internal to the perception modules. While I do not completely rule out this possibility, in what follows I will assume for simplicity that perceptual mechanisms have in their output only perceptual concepts. (The fact that numerical concepts can plausibly occur in perceptual outputs—TWO APPROACHING TIGERS, say—might appear to be an obvious counterexample; but this may say more about our direct perceptual access to small numerosities through subitizing than about the availability of abstract concepts to our perceptual processes.)

an inferential concept in the intended sense. Being non-perceptual and innately given, such concepts can be abstract in a way that perceptual concepts cannot.

A number of different conceptual domains licence spontaneous inferences about the entities relevant to the domain. These domains might include naïve mechanics (Baillargeon 2002), naïve biology (Atran 2002), naïve sociology (Hirschfeld 2001), naïve mathematics (Dehaene 1997), spatial reasoning (Hermer & Spelke 1996), and others. For example, the domain of naïve biology licenses certain inferences about growth, nutrition and reproduction. Sub-domains for animals and plants license more specific inferences. It is not implausible to see these conceptual domains as encapsulated modules and sub-modules in broadly the same way as Fodorian input modules (Sperber 1994, 2005),<sup>39</sup> and the spontaneous inferences as being governed by meaning postulates or procedures governing non-demonstrative spontaneous inference.<sup>40</sup> Note that these modules would differ from perceptual systems in the following sense: rather than taking nonconceptual inputs and giving conceptual outputs, as perceptual systems do, these modules have conceptual representations as both input and output. Their inputs come not from sensory transducers but from other modules (including other conceptual modules). I will refer to such modules, following Sperber (1994), as ‘first-order conceptual modules’; they are to be distinguished from second-order (that is, metarepresentational) modules such as those involved in theory of mind and inferential communication,<sup>41</sup> and from perceptual modules (that is, Fodorian input modules) such as face recognition, voice recognition, and possibly detection of the emotional state of conspecifics (Cosmides & Tooby 2000b).

---

<sup>39</sup> But see Fodor (2000b) for a dissenting view. The arguments that Sperber and others have advanced in support of a modular view of central cognition are critically assessed by Samuels (2005).

<sup>40</sup> When I speak of procedures for spontaneous non-demonstrative inference, I have in mind to leave open the possibility that, in addition to meaning postulates, we may also spontaneously apply probabilistic rules, heuristics, mental models and so on in intuitive reasoning.

<sup>41</sup> Inferential communication (pragmatics) is plausibly a sub-module of theory of mind (see Wilson 2000; Sperber & Wilson 2002).



Here's one fairly standard way to think about this. The hierarchies such as *Blue Point*, *Siamese*, *cat*, *mammal*, *animal*, *physical object* that we looked at above are sometimes referred to in the literature as 'implicational hierarchies', in reference to that fact that membership of a category lower in the hierarchy implies membership of all categories higher in the hierarchy (all cats are mammals, animals and physical objects). This establishes subset or inclusion relationships between lower-level categories and higher-level categories. A second fact about such hierarchies is that entities in lower-level categories inherit the properties associated with higher-level categories. For example, if it's a property of animals that they give birth to young of the same kind, then this will also be a property of cats, and a property of Blue Point Siamese. This second fact follows trivially from the first, since property ascription can itself be seen as just a species of set inclusion: if all *F*s are *G*s, then whatever properties *G*s have as such, *F*s will also have in virtue of being *G*s.

This has obvious ramifications for the organization of knowledge. An individual that has a set of intuitive beliefs about animals (that they tend to seek food when hungry, that they give birth to young of the same kind, and so on) doesn't need to acquire corresponding beliefs on a case-by-case basis about cats, dogs, tigers, agoutis, and so on, provided that they have access to the fact that these are all kinds of animal. Following Sperber (1994) and Sperber & Wilson (1995: Chapter 2), the following picture suggests itself. The categories in mentally-represented implicational hierarchies correspond to intuitive concepts, some general and some more specific. Which intuitive concepts an individual possesses is an empirical question, not one to be settled *a priori* by analytic philosophers, so we cannot expect conceptual hierarchies to map perfectly onto taxonomic hierarchies. A conceptual domain such as naïve biology corresponds to a very general concept such as LIVING KIND, attached to which are the spontaneous inferences that the concept LIVING KIND enters into, as well as a database of intuitive beliefs pertaining to living kinds. As Sperber (1994) points out, such a concept can be seen as a module: it is domain specific, performs special-purpose computations (governed by meaning postulates

and procedures attached to the concept) and has a proprietary database (the encyclopaedic information stored under the concept). More specific living kind concepts (such as ANIMAL, CAT, SIAMESE) are structured in a similar way: they too have attached meaning postulates/procedures and proprietary databases of intuitive beliefs, and they too can be seen as modules or sub-modules. The content of all these concepts is constituted by their nomological links to the corresponding properties in the way proposed by informational semantics. The content is constrained, however, by the meaning postulates attached to these concepts in the way described in chapter 3.<sup>42</sup>

To link this hierarchy of concepts together in a way that is cognitively useful, so that intuitive inferences that can be drawn about animals as such are also drawn about cats, and knowledge about animals as such can be applied to cats, we need the cognitive equivalent of the set-inclusion relation between members of hierarchies. This can take the form of meaning postulates attached to concepts linking them to higher-level concepts in the hierarchy. For example, attached to the concept CAT might be the meaning postulate ' $\varphi_{\text{CAT}} \psi \rightarrow \varphi_{\text{ANIMAL}} \psi$ ' and attached to the concept ANIMAL might be the meaning postulate ' $\varphi_{\text{ANIMAL}} \psi \rightarrow \varphi_{\text{LIVING KIND}} \psi$ '. (I have discussed the status of such meaning postulates in chapter 2.)<sup>43</sup>

It can be seen that conceptual modules perform spontaneous inferences that are governed by the meaning postulates and procedures attached to concepts. These spontaneous inferences can introduce concepts which are non-perceptual—because

---

<sup>42</sup> If in addition to meaning postulates there are procedures for spontaneous non-demonstrative inference in the logical entries of concepts, these too presumably constrain the content of the concepts to which they are attached.

<sup>43</sup> There may appear to be an 'order of acquisition' problem here. Children, as has been noted above, tend to acquire basic-level words such as 'cat' before superordinate-level words such as 'animal'. If this is taken as evidence for the order in which concepts are acquired (cf. Fodor 1981), then the view put forward in the text that conceptual hierarchies are linked together by meaning postulates could be undermined. However, there is no reason to take the lexical acquisition data as demonstrating anything about the order of *conceptual* acquisition. In fact, there is evidence that a number of highly abstract concepts (such as CAUSE and EFFECT or BELIEF and DESIRE) appear long before the corresponding words (Leslie & Keeble 1987; Leslie et al. 2004).

they are concepts for abstract or not perceptually identifiable entities. These are the inferential concepts. To give an example, it may be that we have an animal-detector, which functions on the basis of a number of cues, such as animacy or intentional movement (Premack 1990; Premack & Premack 1995), the presence of eyes (Jones et al. 1991), and so on (see also Gelman et al. 1995), in which case ANIMAL is a perceptual concept. Let us suppose, however, that we do not have a LIVING KIND detector (perhaps we have separate detectors for animals and plants, but no single detector for living kinds in general), in which case LIVING KIND would be non-perceptual.<sup>44</sup> In such a situation, the meaning postulate ‘ $\varphi$  CAT  $\psi \rightarrow \varphi$  ANIMAL  $\psi$ ’ does not introduce any purely inferential concepts (both CAT and ANIMAL are perceptual). In contrast, if LIVING KIND is non-perceptual, then the meaning postulate ‘ $\varphi$  ANIMAL  $\psi \rightarrow \varphi$  LIVING KIND  $\psi$ ’ does introduce an inferential concept. We can thus see how conceptual modules could, through meaning postulates, introduce inferential concepts that are not available in the outputs of perception.

Sperber, recall, suggests that such inferential concepts are “innately pre-formed, unanalysed abstract concepts” (1996: 89). It should now be fairly clear how he arrives at this position. Since such concepts are non-perceptual, it seems that they must be innate; if not, it’s difficult to see how they could be acquired—it would require the ability to learn new module-specific rules of inference, which seems *prima facie* implausible (we will discuss this further in chapter 5). In fact, the only plausible way to have a concept that is neither perceptual nor innate is to construct it from more basic concepts (hence our ability to think PINK UNICORN), but it is unlikely that LIVING KIND in the above meaning postulate is a complex concept in this way. Lastly, inferential concepts will be abstract, almost by definition: whatever we can’t perceive is *ipso facto* abstract in a certain sense (although this is not to make any metaphysical claim: there is nothing in principle to prevent us from

---

<sup>44</sup> For indications that children have difficulty in forming an adult-like LIVING KIND concept, see Wellman & Gelman (1998) and Opfer & Siegler (2004).

acquiring a perceptual detector for living kinds, say, in which case the concept would become perceptual, hence non-abstract in the current sense).

To recap. There are two kinds of intuitive concept: *perceptual concepts* which occur in the outputs of perceptual modules and for which we must therefore have perceptual detectors; and *inferential concepts* which are introduced via spontaneous inferences by conceptual modules. In the next section, I look in more detail at the different kinds of detector that we may possess, and how they operate.

#### **4.4. Different kinds of detector**

We have seen above that perceptual concepts are those for which we have perceptual detectors. It does not necessarily follow that perceptual detectors underlie our abilities to identify all objects: we may identify some objects inferentially rather than perceptually. What is required for possession of a concept is that some internal state of the organism reliably correlates with the presence in the organism's environment of entities falling under the concept (Dretske 1986, 1989); such a correlation does not need to be mediated exclusively by perception. We therefore cannot conclude that we have a perceptual concept for an object purely on the basis that we can identify that object. In this section, I will argue that we have detectors at different levels, including perceptual detectors and detectors at the conceptual level.

How, then, can we tell when an ability to identify a given object is based on a perceptual detector rather than on some conceptual-inferential ability? An important consideration is that perceptual detectors are modality-specific: for example, we may be able to recognize a friend by their face, and also by their voice, but different detectors must underlie these processes. This is clear from architectural considerations—detectors for individual faces will take inputs from visual transducers (indirectly, via early visual processing mechanisms), and will need to perform very different kinds of computations than detectors for individual voices, which will take inputs from auditory transducers (via early auditory processing mechanisms). Evidence of double-dissociations confirms this: prosopagnosics cannot

identify people they know from photographs of their faces, but can identify them over the telephone; phonagnosics show the opposite pattern (see, for example, Neuner & Schweinberger 2000). To identify a familiar person, then, we can make use of at least two distinct detectors, in different sensory modalities, which operate in parallel.

Now, if there are multiple detectors in different sensory modalities that are effectively performing the same task (in this case, identification of familiar individuals), it is not implausible that some supra-modal conceptual process integrates the outputs of these modules. Such a process could help to resolve conflicts between detectors in the identification of an individual and would be particularly useful in facilitating identification in conditions where the stimulus was degraded.<sup>45</sup> Several detailed models that have been proposed to explain how we identify familiar individuals do in fact postulate some process of this kind (Burton et al. 1990, Ellis et al. 1997). We can see such a process as a “conceptual detector”. Importantly, as a conceptual process it could have access to relevant background knowledge, in the form of encyclopaedic information stored under the concept for the individual in question.

Conceptual detectors almost certainly have a role to play in the identification of other kinds of entity, such as living kinds and artefacts. Consider living kinds, in particular animal kinds. Just as we can identify cats by how they look, we can also identify them by how they sound and—at least sometimes—we may need supra-modal integration of this information to allow identification to take place. This suggests that, as is the case with identification of individuals, we have modality-specific perceptual detectors for animal kinds (utilising cues such as shape, texture, characteristic movement, sound, odour), as well as supra-modal conceptual detectors that integrate this information and have access to encyclopaedic knowledge.

---

<sup>45</sup> Cross-modal integration at the conceptual level doesn't rule out the possibility of detectors in different modalities having some influence on each other's operations; there are certain neurophysiological indications that such a possibility may have to be considered (von Kriegstein et al. 2005).

Conceptual detection may be needed when the stimulus is degraded (in poor light, say) or when it is difficult to identify for other reasons. For example, consider the case of an animal that is curled up asleep. In such a situation, shape, movement and sound cues may be unavailable, making perceptual identification difficult, and we may need to rely on conceptual identification via inference: it has whiskers, so it's a cat. Clinical evidence of semantic-knowledge impairment supports the postulation of a supra-modal level of integration specific to living kinds (although in this case the interpretation of the data is somewhat more controversial). For example, there are cases of people who are severely impaired at identifying living kinds, across sensory modalities, but whose capacity to identify nonliving things is largely spared (Warrington & Shallice 1984; De Renzi & Lucchelli 1994; Caramazza & Shelton 1998; Farah 2004: Chapter 9). A number of these patients do not show any impairment in face recognition/person identification, consistent with the view that the supra-modal integration that is occurring is specific to living kinds. These cases of domain-specific semantic impairment lend support to the view that we have a conceptual detector for living kinds.

Next, consider the identification of artefacts. There has been considerable debate in the philosophical and psychological literatures on the question of how we categorize artefacts (Dennett 1990; Bloom 1996, 1998; Malt & Johnson 1998; Matan & Carey 2001). It is fairly clear that for an object to count as being an artefact of a certain kind (a chair, say), the most important consideration is not its shape or even its function. Plenty of chairs have unusual shapes (deckchairs, swivel chairs, massage chairs) and not all objects used in the same way as chairs count as actual chairs (a table used for sitting on is still a table, not a chair). Rather, whether an object counts as being an artefact of a particular kind has more to do with its original intended function. (The philosophical details get notoriously complicated, but this much will suffice for present purposes.) A separate question is how we identify an object as belonging to a particular artefact-kind. And here both shape and function do appear to be relevant. Even if we understand that the essential property for

artefact kind membership is original intended function, this does not mean that we always (or ever) in fact classify artefacts on the basis of this property. Rather, we presumably use other cues that reliably correlate with the essential property, in the same way that we rely on cues of shape, texture, characteristic motion and so on in identifying animal kinds, even if we understand that kind-membership is determined in this case by hidden microstructure. There is some evidence that shape, as well as function—or, rather, the presence of features that reliably correlate with function—are used as a basis for categorizing artefacts (Gelman et al. 1995; Landau et al. 1998b; Booth & Waxman 2002; Jones & Smith 2002). Shape is a reliable, although by no means infallible, perceptual indication of artefact-kind, and it is likely that shape detection plays an important role in kind-identification. As regards function, many artefacts have perceptually-identifiable parts or features that can indicate kind-membership either at the basic or superordinate level (for example, wheels; see Madole & Cohen 1995, Rakison & Butterworth 1998). While the inference from a particular feature to a particular function relies on conceptualized knowledge of the world, it is not impossible that some features are learned as perceptual cues to kind membership (we will look at some examples of learned perceptual cues in chapter 5). These perceptual detectors will be supplemented, and in some cases overridden, by kind-detection at the conceptual level, based on inference about function and creator's intention. Where other information is not available, however, shape and the presence of certain features will be the basis on which inferences about function and creator's intention are made. In the case of familiar exemplars, identification could proceed directly from perception, with no need for inferences about creator's intention—it is implausible that we carry out such inferences each time we see a standard swivel-chair or dining chair, although we may need to in the case of more unusual exemplars (cf. Dennett 1990; Bloom 2000: 162).

The kinds of inferences that are warranted for artefacts are very different from the inferences about living kinds, and it has been assumed that artefacts form a separate, specific domain of knowledge. Clinical evidence again provides some

support for the existence of a supra-modal level of integration that is specific to artefacts. A number of studies have demonstrated impairments of semantic memory that are specific to artefacts—that is, there are cases of people who are severely impaired at identifying artefacts, but whose capacity to identify living kinds is largely spared (Hillis & Caramazza 1991; Sacchett & Humphreys 1992; Farah 2004: 149 f.; see also Caramazza & Shelton 1998). Together with the studies of patients with selective impairments for living kinds, this evidence demonstrates a double-dissociation between the conceptual domains of living things and artefacts, in support of the idea that we have conceptual detectors for artefacts as well as for living kinds.

Animal kind detectors generally operate at the basic level (corresponding approximately to the genus in scientific Linnaean taxonomy), as do detectors for artefact kinds ('chair' and 'car', rather than 'wheel-back' and 'E-type Jaguar'), whereas the detectors underlying the identification of familiar faces obviously operate at the level of individual faces. Face recognition is unusual in several respects. The evidence suggests that there is a perceptual module dedicated to identifying individual faces. This module is highly specialized—it is apparently not recruited, for example, in the identification of inverted human faces, or in identifying faces of sheep (an ability that remained intact in a farmer with severe prosopagnosia; see Farah 2004: Chapter 7). The ability to identify individual faces takes many years to fully develop, and requires exposure to large numbers of faces (Carey 1992).<sup>46</sup> In addition, repeated exposure to a particular face is needed for an ability to easily identify that face to develop: humans are extremely skilled at identifying familiar faces, even when the stimulus is degraded, but are poor at identifying or matching unfamiliar faces (Hancock et al. 2000). The module is not responsible for

---

<sup>46</sup> Note also that a selective set of stimuli leads to selective expertise, which gives rise to the 'other race' effect (which becomes more pronounced with development). See Chance et al. (1982).



categorizing objects as faces, but only for identifying individual faces.<sup>47</sup> This is clear from the fact that prosopagnosics, while unable to identify individual faces, have no difficulty identifying faces *as faces* (Farah 2004: 94), and therefore must have a separate and intact perceptual detector associated with the basic concept FACE. Each perceptual detector which allows us to identify a familiar face can be seen as a sub-module of the face identification module, acquired through experience (cf. Sperber 1994); the corresponding ability to identify entities at the basic level does not require the same level of repeated exposure. We will look at the implications of this for concept acquisition in chapter 5.

We have now seen in more detail how the distinction between the two kinds of intuitive concept—perceptual and inferential—could be cashed out in more detail. Although these details will be important in chapter 5, when we come to look at concept acquisition, for present purposes it is sufficient to note that there is a class of intuitive concepts, and that this class is not limited to perceptual concepts, but extends to inferential concepts not available to perceptual processes. Before moving on, though, here's a couple of refinements.

First refinement. It is important to note that not all perceptual detectors will necessarily give rise to concepts appearing in the outputs of perceptual processes. That is, there is a class of what, following Stich (1978),<sup>48</sup> could be called “subdoxastic detectors”. These detectors give rise to mental representations (concepts, if you like) which can appear in the internal representations of perceptual modules but which, since they are not present in the outputs of these modules, are not available to the organism as a whole. Consider, for example, the detectors for edges employed by low-level vision (Marr 1982). These are important for the internal processes of the visual system, and representations of edges may occur at interlevels of the visual system (say, at the interface between low-level and mid-level

---

<sup>47</sup> Although the module must be sensitive to certain properties of faces as such, in order for faces to trigger the operation of the module. In this respect, non-human primate faces may be similar enough to human faces to trigger the module, but not faces of other species (see Pascalis et al. 2001).

<sup>48</sup> See also Fodor (1983: 83 ff.) for relevant discussion.

vision), but will not appear in the output of the visual system, which is object-based.<sup>49</sup>

The second refinement concerns the supra-modal integration of information. We have been considering the idea that conceptual modules serve to integrate the outputs of different perceptual modules, and that these can also be seen as detectors—they operate automatically and subconsciously, but this time at the conceptual level. But there is also a second kind of supra-modal integration, that of deliberate, reflective thought. Reflective thought has access to all sorts of reflective beliefs and background information that is not available to the limited database of our intuitive inference mechanisms. If perception alone allows us to detect robins that are nearby, and conceptual integration allows us to detect robins that are more distantly glimpsed, then reflective thought permits detection of robins on the basis of information contained in our bird-book.

#### 4.5. Conclusion

We are now in a position to return to where this chapter began, and consider how the distinction between intuitive and reflective concepts can be productively applied to cases of incomplete or poorly understood concepts.

There were two problems, one metaphysical and the other psychological. The metaphysical problem was how to give an atomistic account of concepts for

---

<sup>49</sup> This is not to deny that we can have an intuitive concept *EDGE*. But it need not be that this concept is coextensive with the subdoxastic representation of an edge employed by the low-level visual system; if it (and other similar concepts) were coextensive with the corresponding representations employed by low-level vision, then it would make our investigation of the operations of the visual system far easier. In support of this, consider the case discussed in Fodor (1984c): a standard explanation of the Müller-Lyer illusion is that the configuration of one set of lines is interpreted as an edge receding from the plane and the other as an edge protruding from the plane, although we are not aware of this fact, and do not even see the configurations consciously as three-dimensional representations. And note, by the way, that it's not just detectors for low- and mid-level sensory properties (line, edge, local contour and so on) that might be subdoxastic. It's not implausible that we also have a subdoxastic visual detector for *face* (to identify a stimulus for processing by the face module) or perhaps *animal* and *artefact* (if we have visual modules specific to animate and inanimate objects; cf. Rakison & Butterworth 1998).

nomologically impossible entities such as GHOST. Recall that since there can't be laws about ghosts, informational atomism can't account for the content of GHOST unless it turns out that this is a complex concept constructed from concepts for possible entities. We possess large numbers of concepts for nomologically impossible entities, and it would be preferable not to have to claim that these are all complex. The psychological problem was that while deference to experts can provide an answer to the metaphysical question of how a concept gets its content, it doesn't explain how we are able to think with such concepts.

Sperber's distinction between intuitive and reflective concepts provides an answer to the psychological problem. Take the example of the concept ECHIDNA that we discussed at the start of this chapter. We imagined someone with no experience of echidnas or echidna-representations, but who knew certain facts about echidnas (that they are a kind of animal from Australia, say), and who therefore must possess some sort of ECHIDNA concept. There is no metaphysical problem in this case, since the person could be locked to the property of *being an echidna* via deference to experts. But this is unsatisfying from a psychological point of view since it leaves us with no account of how we actually think with such concepts. If we adopt Sperber's intuitive-reflective distinction, however, we can propose that for the individual in question the concept ECHIDNA is reflective. They can think reflectively about echidnas, using the little information that they have gleaned about them. Their knowledge that echidnas are a kind of animal will also enable them to reflectively draw inferences about echidnas *qua* animal. The person therefore has a reflective ECHIDNA concept, but they have no perceptual detector for echidnas. It may nevertheless still be possible for this person to use the background information that they have about echidnas to (tentatively) identify one. That is, the person's encyclopaedic information about echidnas may be sufficient in some contexts to enable them to infer the presence of an echidna; a disposition to defer to others will ensure that any tentative echidna-identification can be overridden. Of course, acquiring further information about echidnas or schmoos (such as encountering some

of the former, or the discovery of more complete specimens of the latter) would enable our reflective concepts in each case to become intuitive; concept acquisition presumably often proceeds in this way.

Next consider the metaphysical problem, which arises for concepts such as GHOST. Since the property of *being a ghost* is not nomologically possible, informational semantics has no resources to account for the content of such concepts, other than by stipulating that they are non-atomic (that is, phrasal). Now, concepts like GHOST are clearly reflective in Sperber's terms: we have no perceptual detectors for ghosts, and GHOST is surely not introduced by any meaning postulate that we employ. In fact, it seems that in a certain sense such concepts have a defective metaphysics, since in point of nomological necessity we cannot be locked to the corresponding property. I would suggest that concepts for nomologically impossible entities, like GHOST, are properly seen as *not having any externalist semantics* (we will consider another case like this below).<sup>50</sup> In this case, there is no need to propose that GHOST is complex—rather, we can just say that it is a reflective concept with a permanent metaphysical defect. The fact that it is metaphysically defective will not prevent the concept from being deployed in (reflective) thought. It can have a lexical entry associated with it, as well encyclopaedic information stored under the concept.

Note that it is certainly not the case that all reflective concepts are metaphysically defective. For example, take the concept PROTON. We don't have the ability to reason intuitively about protons, or perceptual detectors for these entities, so PROTON—like GHOST—is a reflective concept. But there is (so physicists assure us) a property of *being a proton*, and so the concept PROTON can get its content in the normal way, by locking to the relevant property—that is, via a theory of protons in the case of experts.

---

<sup>50</sup> Cf. Segal (2000), who argues that empty concepts require an internalist treatment, and takes this to support an internalist account of content more generally. I am in a better position to resist the move to internalism for non-empty concepts, since I can insist that only reflective concepts can be metaphysically defective and hence in need of an internalist semantics.

Just as certain concepts (like GHOST) will be permanently reflective because of a metaphysical defect, certain other concepts may be permanently reflective for other reasons. Consider the so-called “attributive use” of concepts discussed by Sperber & Wilson (1983: Chapter 8; 1995: Chapter 4). Take the following example that they discuss. Presumably the concept CLAIRVOYANT is uninstantiated: there are in fact no people who are able to remotely perceive objects or events. And presumably, many people believe that there are no clairvoyants. For these people, use of the word ‘clairvoyant’ and hence of the concept CLAIRVOYANT would be attributive rather than descriptive.<sup>51</sup> Suppose I utter the sentence in (11):

(11) The use of clairvoyants by police is a waste of taxpayers’ money

The word ‘clairvoyant’ here (and the corresponding concept) means something like “people who claim or are claimed to be clairvoyant”. Its use is attributive since uttering (11) need involve no commitment to the existence of actual clairvoyants, merely to the existence of individuals who are attributed with such abilities by themselves or others. As Sperber & Wilson point out, deference to experts involves *endorsing* whatever the content of the expert’s concept is. A reflective concept, however, may or may not involve an endorsement of the content.<sup>52</sup>

---

<sup>51</sup> The same point can be made about GHOST, of course. In the above discussion of GHOST, however, we were assuming that the individual in question did believe in ghosts, and therefore used GHOST descriptively rather than attributively.

<sup>52</sup> This approach provides a response to Cain’s (2002: 77) worry about the concept CHIC. Cain notes that we may require considerable immersion in popular fashion culture to acquire the ability to identify chic individuals, and he suggests that this undermines Fodor’s account of the possession of mind-dependent concepts. First, though, it’s perfectly okay for Fodor if acquiring the concept X in practice requires acquiring lots of other concepts; what Fodor can’t allow is that *possession* of concept X requires possession of other concepts. Second, on the present framework the concept CHIC would be reflective for most of us, in which case we would not be expected to be able to identify chic individuals as such, either in the case where we use the concept descriptively, or if we use the concept attributively.

## 5. Acquisition

### 5.1. Learning instincts

There was once a fairly neat dichotomy between instinct on the one hand and learning on the other. Ethologists saw instinct at work in their observations of animal behaviour in natural settings (Tinbergen 1951). Nest-building by birds and web-spinning by spiders were clearly innately-determined behaviours rather than learned skills. More generally, each individual species could be seen to possess a range of typical behaviours which were often expressed even by individuals raised in captivity in isolation from conspecifics.

By contrast, behaviourist psychologists studying learning in the laboratory considered that complex behaviours arose from associative learning via processes of conditioning (Watson 1925). It was believed that animals could learn to associate arbitrary stimuli with arbitrary behaviours. Behaviourists came to completely eschew appeals to instinct, arguing that all behaviour should be explained in terms of conditioning. This is summed up by Watson's statement: "...try hardest of all to think of each unlearned act as becoming conditioned shortly after birth—even our respiration and circulation".<sup>1</sup>

Traditionally, two different forms of conditioning were distinguished: classical conditioning and operant conditioning. Classical conditioning was first described by Pavlov in his work with dogs (see Pavlov 1927). He showed that animals were sensitive to innately-recognized cues (food, say) which they responded to with innately-specified behaviours (salivation, say). If a novel stimulus (a ringing bell, say) was consistently presented a few seconds prior to the cue, then over time this novel stimulus alone would come to elicit the behavioural response (that is, Pavlov's dogs would salivate when the bell was rung even in the absence of food). In the other form of conditioning, operant conditioning, animals learn novel complex behaviours

---

<sup>1</sup> Watson (1925: 105), cited in Gould (2002).

(pressing a lever, say) by trial and error, with feedback from the effects of the action (a reward, say) shaping the behaviour.

As the experimental data accumulated, behaviourism began to face serious difficulties (see Gould 2002 for discussion). It started to become clear that some species could learn certain associations more readily than others. For example, rats can readily learn to associate an olfactory stimulus with a food reward and an auditory stimulus with a shock; but they are very slow to associate an odour with a shock or a sound with a food reward. Various asymmetries of this kind are found in other species. These facts undermined the claim that animals do not have any innate predispositions to form particular associations in particular contexts. And even the basic principles of conditioning were called into question. Experiments had shown that rats can normally link cues with effects—pressing a lever to avoid a painful shock, as it might be—only when the delay between the two is very short (no more than a few seconds); but despite this, rats that become sick were shown to develop aversions to foods consumed several hours earlier (Garcia & Koelling 1966).

Even after it had become clear that there were both innate and environmental factors at play, however, the dichotomy persisted. The debate over whether behaviour was to be explained by instinct or conditioning was supplanted by a debate over whether *particular* behaviours or traits were to be considered innate or learned. This was no doubt fuelled by loose talk of genes for a wide range of traits (a “gene for Alzheimer’s” or a “gene for homosexuality”; see de Waal 1999a). Thus, there were some behaviours that were pretty clearly innate, such as web spinning in spiders (there is very little variability across individuals, and no learning is apparently required; see Reed et al. 1970) and song production in some species of birds (for example, North American flycatchers, where song develops normally in birds raised in isolation; see Marler 2004). Then, there were other behaviours that were pretty clearly learned, such as song production in certain different species of birds (for example, sparrows, which need to learn their songs from ‘tutors’; again, see Marler 2004).

Something of the same dichotomy was evident in discussions of concept learning, as Fodor has discussed. He noted that whatever account of learning one might adopt, concept learning had to consist of some process of hypothesis formation and confirmation; acquiring a concept as the result of a blow to the head or from taking a drug just wouldn't count as concept *learning* (Fodor 1975, 1980; see also 2001a: §3).<sup>2</sup> Now, hypothesis formation and confirmation is pretty much an analogue to the kind of trial-and-error learning involved in operant conditioning.<sup>3</sup> And very often 'concept-learning' experiments of the time were couched in more-or-less the same behaviourist vocabulary of stimulus, response, feedback and so on. But as Fodor (1975, 1981) famously pointed out, hypothesis formation and confirmation only works as an account of how complex concepts are learned; primitive concepts cannot be learned in this way, since one would already need to possess the concept in question in order to frame the relevant hypothesis in the first place. Combine this with an argument that most (lexical) concepts are primitive and you reach the (in)famous conclusion of Fodor (1981) that most (lexical) concepts must be innate. We seem to have a similar dichotomy here to the one found in the case of behaviour:

---

<sup>2</sup> For discussion of some naturalistic cases of non-psychological processes which have specific psychological effects, see Samuels (2002: §5). For example, Samuels discusses the case of Ross River Fever, a tropical disease which is said to cause its victims to experience very specific hallucinations (buildings falling down on the road in front of them). One might add other examples of this kind. Thus, rabies has evolved to cause aggression in its hosts (a very specific psychological effect), plausibly to enable it to jump to another host when saliva gets into the wound. An even more striking example is that of the parasite *Toxoplasma gondii*. The life-cycle of this parasite requires it to infect an animal host, usually a rodent (in order to mature) and subsequently a cat (the only animal in which it can sexually reproduce). To ensure that the rodents it infects do get eaten by cats, the parasite has evolved to have the (incredibly specific) psychological effect of inducing infected rodents to lose their inbuilt aversion to, and fear of, cat pheromones. The rodents are otherwise unaffected (for example, they have normal reactions to the pheromones of the opposite sex). (See Sapolsky 2003 and Berdoy et al. 2000 for discussion of these last two cases).

<sup>3</sup> Although, of course, behaviourists themselves saw concept learning not as an *analogue to*, but rather as a *kind of*, operant conditioning. That is, what they considered to be going on was just another kind of behavioural modification—learning a behavioural response to a particular stimulus, modulated by feedback, just like rats learning to press a bar. Cf. Fodor (1975: 35, fn. 6).



either concepts are acquired through a trial-and-error learning process (in this case, hypothesis formation and confirmation), or otherwise they are innate.<sup>4</sup>

What has emerged over the last couple of decades from the debate over innate versus learned behaviour is a more nuanced understanding that there is a complex interplay between innate and learned, instinct and environment. This is captured by the notion of “instincts to learn”, developed by Gould and Marler.<sup>5</sup> Take as an example the question of song learning in birds. Since some bird species produce their species-typical song even when raised in isolation, whereas other bird species will not, it appeared that song production was innate in some species and learned in others. But detailed studies (such as those reported in Marler 1991) showed that the picture was actually far more complex. While it was clear that species with “learned” song did produce highly aberrant song when raised in isolation, detailed examination of these songs nevertheless revealed that they exhibited a number of species-typical characteristics (as regards structure and duration, for example). Even the classic example of web spinning in spiders has recently been revealed to be far more complex than originally thought. Heiling & Herberstein (1999) show that experience plays an important role in web spinning. Experienced web-spinners construct webs that are more asymmetric (a feature that enhances the efficiency of prey-capture), and both experienced and inexperienced web spinners adjust the degree of asymmetry in accordance with prey-capture statistics over preceding days.

The moral is that a particular behaviour is better viewed not as being innate or learned, or even as being more innate or more learned, but rather as arising from the

---

<sup>4</sup> Fodor had in mind a fairly restrictive sense of ‘learning’ as a rational process mediated by (intentional) psychological mechanisms. The notion of ‘instincts to learn’ involves a less restrictive sense of ‘learning’ as acquisition by some psychological (but not necessarily intentional) process. It is generally the latter—more general—sense of ‘learning’ that I use in this chapter. For detailed discussion of how the terms ‘innateness’ and ‘learning’ have been employed in cognitive science, see Samuels (2002).

<sup>5</sup> See Gould & Marler (1984, 1987); Gould (2002); Marler (1991, 2004); see also de Waal (1999a). For discussion of the difficulties associated with drawing a principled innate/learned distinction, and some proposals for how best to formulate the notion of ‘innateness’, see Samuels (2002, 2004); see also Fodor (2003b).

complex interplay of innate and environmental influences on the course of development and learning. Such instincts to learn “set a species-specific context within which experience operates” (Marler 1991: 59). It is very clear why an organism would benefit from learning instincts. In any learning situation, there is an indefinite number of possible cues to which the organism could attend. Without instinctive biases to attend to certain cues rather than others in particular situations, it would be practically impossible to learn anything at all. The stronger and more situation-specific the relevant biases can be, the easier the learning task for the organism. But, of course, instinctive biases come at a cost of reduced flexibility. Environments have a nasty habit of changing, and cues that once had high ecological validity can become suddenly much less useful. Mother nature must therefore balance the relative advantages of innate biases towards those cues that are stable and adaptively significant, and learning of important but contingent aspects of ever-changing environments.

Take, for example, the task of food/prey selection (see Gould 2002). Specialist organisms consuming highly specific diets can be more reliant on innate cues for identification of food/prey. For example, digger wasps prey on a single species, and their stinging behaviour is highly specialized for the anatomy of this particular species. Unsurprisingly, prey selection in these wasps is based on highly specific unlearned cues, and thus innately constrained to their single prey species. A very different approach is needed in the case of generalist feeders. Rats, for example, require a general strategy for identifying the range of safe and acceptable foods available in the local environment, rather than cues to identify a specific prey species or specific foods. Studies have shown that rats learn to identify novel foods on the basis of odours on the breath and whiskers of conspecifics.

One might imagine that these differences between wasps and rats could relate to the fact that learning requires more complex cognitive machinery, and therefore that simple organisms will tend to have simple, innately-specified behaviours. But this is not the case. Bees have a sophisticated ability to learn the identification and location

of nearby nectar-producing flowers, and even the time of day when flowers of particular species are most productive. The cues to which the bees are sensitive, however, are governed by innate biases; for example, bees make use—in descending order of importance—of odour, colour and shape cues in identifying flowers, but they are unable to make use of information about light-polarization, even though this is something which they are highly sensitive to (Gould & Marler 1987). Digger wasps themselves have been shown to be extremely effective learners in contexts other than prey selection—for example, they are able to learn the locations of multiple nests and the condition of their offspring in each nest (Gould 2002). More generally, even complex organisms need to take the best advantage of reliable features of the environment so that they are able to maximise cognitive efficiency; what might appear to be relatively simple tasks could otherwise easily become computationally intractable. For example, incest-avoidance in humans and other primates is mediated by a relatively simple mechanism whereby early familiarity—such as would normally occur between siblings, or between offspring and parents—kills sexual desire (the so-called ‘Westermarck effect’; see de Waal 1999a). This is an example of a very simple mechanism that produces an adaptive effect by approximating a desired outcome, in this case avoiding the negative consequences of inbreeding. It exhibits an interplay between learning and instinct: the consequences of early familiarity are innately prescribed, but the particular set of individuals to whom these consequences apply is acquired.

## **5.2. Instincts to acquire concepts**

The literature on learning instincts tends to show that the innate–learned dichotomy breaks down as a characterization of behaviours. This raises the question of whether the dichotomy might not also break down in the case of *ideas*. If so, then it could suggest possibilities for escaping from Fodor’s (1981) logic and avoiding radical concept nativism. (This is a direction that Fodor himself has taken in recent work, as we saw in chapter 1.)

The usual starting point in contemporary discussions of concept nativism is with the empiricist assumption that other than a relatively small number of sensory primitives and logico-syntactic terms, all lexical concepts are learned. An attempt is then made to explain how it is that these concepts can be learned (by providing definitions, say, or partial decompositions, or prototype structures). Any concept that one's chosen theory has fundamental difficulties with may then be allowed into the innate base—so, for example, Wierzbicka's 'natural semantic metalanguage' started out with 14 innate 'primes' in the 1970s (Wierzbicka 1972), and now has more than 60 (Goddard & Wierzbicka 2002).<sup>6</sup>

Given, however, Fodor's powerful arguments that approximately *no* lexical concepts can be learned, let us instead take as a starting point the opposite assumption that essentially all lexical concepts are innate, and see what difficulties such a view might face. Now, it may be plausible that evolution has endowed us with a certain number of innate concepts. The reasoning is more-or-less parallel to the case of innate behaviours: innate instructions are more efficient and less error-prone than trial-and-error learning, so important but predictable tasks will tend to be innately-governed (nest-building, say), with learning playing a more minor role (fine-tuning the relevant motor skills, say). This argument is particularly strong in the case of behaviours where there is little room for error, in particular behaviours having life-or-death consequences for the organism, such as avoiding vertical drops. In the same way, it is not implausible that evolution has endowed us with certain innate concepts, by means of innate detectors which just need to be triggered by the right environmental input. This could be the case with SNAKE, for example, as well as concepts for a number of other specific dangers and more general distinctions (see §4.1.3 above and Tooby et al. 2005, Sowards & Sowards 2002). The role of learning would then be to fine-tune the relevant detectors based on experience of the entities

---

<sup>6</sup> Granted that if we extrapolate from the historical trend it will still be some years before natural semantic metalanguage includes all eleven-thousand-odd monomorphemic words in the CELEX database of English words (Baayen et al. 1995).

in question. Given that the whole rationale for evolution to have innately endowed an organism with these concepts would be their key importance to the survival of that organism, it would be important for the innate detectors to be set up in such a way that the number of false negatives is minimized. They are therefore likely to be based on fairly straightforward cues that will produce few false negatives but probably quite a few false positives. The innate component of a snake-detector may consist in nothing much more complicated than the kinds of cues, for example, that toads are sensitive to in detecting worms: elongated objects of a certain size moving in the direction of their longer axis (Wachowitz & Ewert 1996). One role of learning would be to fine-tune these detectors in order to reduce the number of false positives while keeping the number of false negatives low.<sup>7</sup> In other cases there is evidence that an innate detector is replaced by a learned detector, or in some cases a series of more specific learned detectors, based on different cues (see Gould 2002: 240, Sowards & Sowards 2002).

While it is plausible that many behaviours are innately-governed, it just cannot be the case that they all are (Gould & Marler 1984, Gould 2002). Bee navigation on the basis of local landmarks depends on knowing what those landmarks are—something which just has to be learned, even if this learning is itself directed by strong innate biases; where there is unpredictability there must be learning. In other cases, there will be an opportunity over time for learning to fine-tune innately-governed behaviours to take account of local environmental conditions. Thus, spiders are able to improve certain (limited) aspects of orb web construction by learning what works best in their own particular situation.

In the same way, it just cannot be the case that evolution could have endowed us with innate detectors corresponding to *all* lexical concepts. Artefacts, in particular, can be designed for an unlimited number of more-or-less arbitrary functions, and can take an unlimited range of forms. We cannot come pre-equipped with detectors for

---

<sup>7</sup> Other roles for learning might include fine tuning of the innate behavioural response.

all artefacts we could potentially encounter, any more than bees could come pre-equipped with knowledge of the locations of all the local nectar sources, or rats could come pre-equipped with knowledge of all suitable foods they might encounter. To the extent that radical concept nativism is plausible at all, it can't be based solely or even principally on innate detectors.<sup>8</sup> In fact, it is *because* radical concept nativism cannot assume innate detectors in the general case that the DOORKNOB/doorknob problem<sup>9</sup> arises in the first place. If the innateness of DOORKNOB, say, was explained by an innate doorknob-detector, then it would be no mystery why it was typically experience of doorknobs that triggered DOORKNOB. This of course raises the question of what innate concepts could possibly be; maybe something dispositional (see Cowie 1999: §4.4 and Fodor 2001a: 136 ff.).

So if Fodor is right that primitive concepts can't be learned, and that most lexical concepts are primitive, then the only way to avoid the conclusion that most lexical concepts are innate would be to develop some plausible account of how primitive concepts could be both unlearned and not innate. That is, we need some account of how concepts can be acquired through a process other than hypothesis-formation-and-confirmation learning. The literature on learning instincts can provide clues as to the form such an account might take (cf. Fodor 2003b).

Ideas, like behaviours, can be seen as arising from the complex interplay between instinct and environment. As in the case of behaviour, then, it may not be particularly relevant to ask of a particular concept whether it is innate or learned. This being said, it can still be useful for expository purposes to identify a range of different cases. So, there are clearly some cases of behaviour where the innate constraints are stronger (like web spinning), and the same can be said of concepts for

---

<sup>8</sup> Famously, the innate immune system is able to respond to virtually any antigen on the basis of a vast innately-determined set of distinct 'detectors' (that is, antibodies), without any process of 'learning' (see Piattelli-Palmarini 1989). This is indeed striking and impressive, but the range of possible antigens is nevertheless constrained by biology in a way that, say, the range of possible artefacts is not.

<sup>9</sup> For discussion of the problem, see §1.2.6 above, and Fodor (1998a: Chapter 6; 2001a).

which the corresponding detector is innate (like SNAKE). Then, there are cases of behaviours governed by an innate template, but where there is a significant element of learning unpredictable environmental variation (as when bees learn the locations of local landmarks, or some species of birds learn songs from conspecifics). For concepts, we will see evidence in the following sections that a very similar mechanism applies in some cases—that is, an innate mechanism governing the acquisition of category-specific cues (something like learning the relevant prototype, as suggested by Fodor 2001a). Lastly, there are cases where a novel behaviour is fully learned, without the aid of innate biases. Such cases are probably quite rare, since it appears that most animals cannot learn new behaviours from observation (Galef 1988). One exception might be “cultural” behaviours in some apes (de Waal 1999b, Whiten et al. 1999, Whiten 2000). For concepts, there is evidence that in some cases we are not innately biased towards the salient cues, and in these cases the acquisition of a detector requires teaching, or at least a good deal of motivated self-study. This is the case, for example, in some expert domains—the subject which we turn to next.

It is ultimately a matter of stipulation whether one wishes to see processes such as the last two as a kind of learning, or whether one prefers to insist (with Fodor 1975, 1981, 2001a) that concept learning must be mediated by hypothesis formation and confirmation—in which case these processes would count as concept acquisition but not concept learning.

### **5.3. Quotidian categorization and expert categorization**

#### ***5.3.1. Expert abilities***

There are various groups of experts who have abilities that non-experts find striking and difficult to comprehend. Expert chicken sexers are able to rapidly determine the sex of day-old chicks based on a visual inspection of their genitalia without necessarily knowing the basis for their decision (see below); expert chess players can

just ‘see’ what their next move should be without needing to evaluate various positions (Simon & Chase 1973); expert wine tasters have the uncanny ability to identify wines and vintages (James 1890: Chapter 13; Solomon 1990; Hugson & Boakes 2001); medical experts can diagnose diseases on the basis of subtle information (Patel et al. 1994); highly experienced interrogators have the uncanny ability to detect lies on the basis of “micromomentary” facial expressions (Ekman et al. 1999); and there are many other examples from various fields (see Ericsson & Lehmann 1996 for a review of the relevant psychological literature).

These expert abilities are hard-earned. In general, it takes several years of regular and deliberate practice to develop these skills to expert level. And in most cases, the basis for such skills—for example, the cues being utilized to make perceptual discriminations—are not accessible to introspection. The possibility of having rapid feedback is important for improving the accuracy of such discriminations. (See Ericsson & Lehmann 1996 for further discussion.) In what follows, we will look in more detail at these features of expert performance in a number of domains.

### ***5.3.2. Chicken sexing and plane spotting***

Take the example of chicken sexing. A charming and detailed account of the history and practice of this arcane art is given in Martin (1994); see also Masui & Hashimoto (1933) for the first detailed account in English of the method, written by the scientists who developed it, and Lunn (1948) for a concise early account. For commercial poultry breeders it is important to identify the female chicks as early as possible, to avoid unnecessary feeding of unproductive male chicks (male birds produce no eggs, of course, and so are of no use to egg producers; they also produce lower-quality meat in most species, and can be very disruptive if raised together with females). Poultry breeders once had to wait until chicks were five to six weeks old before they could differentiate male from female, on the basis of clear differences in the adult feathers which start to appear at that time. Then, after several years of intensive research in the 1920s three Japanese vets discovered a method for sexing



day-old chicks on the basis of subtle differences in genital (or “vent”) configuration. It then took several years for the first experts to develop commercially-useful levels of speed and accuracy.

Chicken sexing exhibits a number of features that are typical of such expert perceptual skills:

*It is a difficult skill to acquire.* Even after some of the basic morphological distinctions were understood, the difficulty of making the relevant perceptual discriminations with any useful degree of speed and accuracy limited the practical application of the technique. One industry expert was quoted as saying in 1930 that “sexing young chicks is not sufficiently accurate to be [of] economic importance to the industry because it allows too great a pullet error, it takes too long, and is a difficult and complicated method to learn” (cited in Martin 1994: 8). A scientific article on the subject in 1948 noted that chick-sexing schools had been established in the United States, but “the task at first seemed impossible to those who attempted the job” (Lunn 1948: 281).

*It requires explicit training.* Expertise in chick sexing does not develop spontaneously from examining chick genitalia. Developing the skill requires (in addition to dexterity and good eyesight) several months of explicit training followed by years of practice to achieve expert speeds of over 1000 chicks per hour at close to 100 per cent accuracy. The training consists of demonstrations of the various different genital configurations (by way of diagrams or photographs)<sup>10</sup> and/or

---

<sup>10</sup> Biederman & Shiffrar (1987: 642, 643) note that the expert sexers they used as subjects did not recall ever having been shown diagrams or photographs of chick genitalia in the course of their training. They also indicate that, to their knowledge, the University of Minnesota photographs of type-variations in chick genitalia that they used in their experiment (which were the same photographs that appeared in Lunn 1948: 286) are the only ones in existence. However, Martin (1994: 232 ff.) reproduces several different sets of photographs and diagrams of type-variations used in chick-sexing schools over the years, and he underlines that “Almost without exception, the most skilled chick sexers I interviewed during the course of this study had sketches, photos and their own drawings of the various types...” (1994: 221). He goes on to note that “A student attending class, or being taught privately, will be shown diagrams or photos of some of these different types” (1994: 223). Like the exotic tales of Eskimo snow discrimination, it seems that the bizarreness of chicken sexing encourages mythologizing, since the misnomer that there is no rational basis for chicken sexing is oft

practice sessions using live chicks with feedback from expert instructors as regards accuracy (Martin 1994: 221 ff.).<sup>11</sup> Particular attention is paid to configurations that are rare or confusing, and some relevant diagnostic cues for these are explicitly pointed out (Masui & Hashimoto 1933; Martin 1994: 223–224). Over the years, as teaching methods improved, the minimum levels of accuracy considered acceptable by the industry increased (Martin 1994: 227). Of course, the first generation of chick sexers did not receive explicit training. They taught themselves over several years of careful experimentation and practice and had to conduct post-mortems on chicks in order to check the accuracy of their discriminations (Martin 1994: 11, 50).

*Discriminations are often intuitive.* Expert chick sexers report that in many cases they are unaware of the basis on which they make their perceptual discriminations, as the following comments illustrate:

In the course of research for this book and listening to conversations between “old time” (vent) chick sexers, one still sexing commercially at 66 said that “Some of the cockerels have nothing there, but I know they are cockerels”. This is intuition at work. He was not conscious of anything that showed it was a cockerel, yet he knew it was. (Martin 1994: 42)

Ben Salewski...claimed that he could sense the sex of the chick by touch as he everted the vent, and no doubt there are other experts who have felt and used similar intuition in sexing. (Lunn 1948: 287)

“When I sat for [the 1937 government chick-sexing examinations] I just looked at the vent. It either had a cockerel eminence or not, and that was it. I was like an athlete, I was at my peak, I had sexed over a million chickens.” —Frank Evans, renowned Australian chick sexer (quoted in Martin 1994: 116)

---

repeated (for example, Dreyfus 1990, Musgrave 1993: 7; see Pullum 1991 for a sober evaluation of the Eskimo snow case).

<sup>11</sup> Obviously the instructors themselves must have close to 100 per cent accuracy in order to give the correct feedback (Martin 1994: 18).

This is something that has attracted the attention of both philosophers (such as Brandom 1998) and psychologists (such as Harnad 1996; see also Biederman & Shiffrar 1987). Philosophers find this of interest because it appears that chick sexing is a case of knowledge in the absence of (rational) justification. Psychologists are interested in part because chick sexing is a capacity which, although it is learned, is apparently not accessible to introspection.

Many expert perceptual skills exhibit similar features to chick sexing (see E. J. Gibson 1969 for discussion). Consider the case of soldiers who must learn to distinguish friend from foe in battle. For obvious reasons, such discriminations need to be fast and accurate. In addition, a large number of sometimes very similar stimuli need to be discriminated. The task is made more difficult by the fact that the stimulus will often be degraded (a tank glimpsed briefly through trees; a fast-moving aircraft seen in the distance).

The need for rapid aircraft identification first became vital during the second world war. As Allan (1958) discusses, formal training was first given in Britain in 1940, prompted by the threat of imminent invasion. Expert 'spotters'—people who had a high level of skill in aircraft differentiation—did exist, but there were too few of them. Training centres were therefore set up, but the problem was that the experts had no idea how they had acquired their skills in the first place, or how to transmit those skills to others (Allan 1958: 246). The expert spotters were aviation enthusiasts whose recognition skill was not the result of training, but had apparently emerged from their general interest in aircraft. Training regimens therefore had to be developed somewhat by trial-and-error.

A number of different approaches have been tried. The first formal training given by the British armed forces in 1940 immersed students in the technical aspects of aircraft design, with recognition skills developing as a by-product of this more general study, as it had for the early self-taught experts. The different physical features of aircraft came to be learned indirectly, through an understanding of basic principles of aviation and aircraft design. The method was not particularly efficient,

and it was difficult for trainees who were not already interested in the subject (Allan 1958: 246). At around the same time, the United States armed forces developed the so-called 'WEFT system' (an acronym for 'wings, engines, fuselage and tail'), which encouraged systematic attention to specific details, but without giving the detailed technical background. This system is the one in use today by most armed forces. Trainers are encouraged to give specific instruction as to the most important distinguishing features, together with pairwise comparison of similar aircraft (see United States Army 1996). The same is true of training in the recognition of other kinds of threat, such as tanks (see Biederman & Shiffrar 1987). As with chicken sexing, rapid feedback is important in ensuring accuracy, and 'flash cards' (with pictures of an aircraft on one side, and the name of the aircraft on the reverse) are often provided to soldiers for self-study.

In contrast to approaches based on distinguishing features, a number of psychologists experimented with a rather different system for teaching aircraft recognition, based on very brief tachistoscopic presentation of pictures of aircraft in real-world settings. Such an approach was inspired by Gestalt psychology, and was known as the 'Renshaw system' (after Samuel Renshaw, the Ohio State University psychologist who introduced it). The system was adopted briefly by the US Navy and Air Force, but later dropped when it became clear that, although trainees did develop the ability to recognize rapidly flashed pictures of planes, it was not particularly effective at developing the most important skill of accurate recognition at maximum distance (see Allan 1958).

A somewhat related method, known as the 'Sargeant system', was inspired by J. J. Gibson's 'ecological approach' to perception (Gibson 1966, 1979). According to Gibson, who was himself influenced to a certain degree by Gestalt psychology, perception does not rely on mental processing or inference, but on the direct selection of information present in the ambient light.<sup>12</sup> On this approach, object

---

<sup>12</sup> See Fodor & Pylyshyn (1981) for a critical discussion of Gibson's approach.

recognition is a matter of directly “picking up” certain “invariant properties” of the object which are present in the ambient light. We are able to learn which particular properties or features of an object are invariant because we have an inbuilt propensity to notice regularities in the stimulus flux (cf. E. J. Gibson 1969: 465). Accordingly, the Sargeant system of recognition training emphasised discovery of distinguishing features by the trainees themselves, without any formal instruction. “The function of the instructor is not a teaching one. The pupils virtually teach themselves...” (Allan 1958: 248). All that was required was that the trainees be provided with photographs of aircraft “well printed on good paper” (Allan 1958: 247). This approach was tried by the British RAF in the 1950s but was found to be ineffective and was subsequently dropped. One interesting observation, however, was that students trained in this way were able to recognise aircraft without consciously recognizing their various different parts (Allan 1958: 251).

Now, if it is correct (as Gibson claims) that we have some innate capacity for perceptual organization, then this could explain why it is that students trained in the Sergeant system develop aircraft recognition skills that are subconscious, in the sense that they are not aware of the basis on which they recognise a particular kind of aircraft as such. The way we learn perceptual skills such as aircraft recognition would then be no different to the way we acquire the categorization skills for everyday objects, which obviously requires no explicit training or learning of features. A Gibsonian would certainly be pleased if this turned out to be the case.<sup>13</sup>

On reflection, though, it's not clear that it does turn out to be the case. One possibility is of course that the intuitive nature of the skill is due to the fact that the knowledge was acquired implicitly and is not available to introspection for that reason. Allan (1958) makes this claim in the case of aircraft recognition.<sup>14</sup> The

---

<sup>13</sup> Interestingly, it was Gibson himself who was charged with preparing a research report on recognition training for the US armed forces, in which he was critical of both the WEFT system and, particularly, of the Renshaw system (Gibson 1948; see also Gibson & Gagné 1947).

<sup>14</sup> But notice that Allan has a problem. On the one hand, introspection gives the impression that we don't make use of features in the course of our everyday categorizations. This was part of the

problem, though, is that these skills *don't* develop automatically in the same way as everyday object categorization. This is precisely what the failure of the Sargeant aircraft-recognition system demonstrates. More generally, there is considerable evidence (reviewed below) that, in contrast to everyday categorization, the development of perceptual expertise requires explicit training and deliberate practice. This suggests that everyday classification and expert recognition are acquired via different processes.

So, the difficulty that untrained subjects have with tasks such as aircraft identification is likely due to the fact that the special-purpose mechanisms which evolved for similar tasks (face recognition, animal kind classification, and so on) are not of much help in these cases. To improve speed and accuracy, current instruction methods involve explicit training focussing attention on certain highly diagnostic cues. Notice that this is different from the chicken sexing case, where untrained subjects have no discriminatory ability, since there are no obvious differences between male and female chicks.<sup>15</sup> But in both cases training is most efficient if particular diagnostic features are explicitly pointed out, and feedback is provided on the accuracy of identifications.

A final case I want to consider is that of bird-watching. This presents similar challenges to tank/aircraft identification, though for somewhat different reasons. Bird-watchers need to make quick and accurate identifications on the basis of often impoverished information. Many birds look alike, at least to the untrained eye (many

---

inspiration for Gestalt psychology, and also influenced Gibson. On the other hand, if perceptual recognition is governed by innate—and subconscious—capacities for picking up certain features of the stimulus, as Gibsonians claim, then this could explain why we are not aware of making use of features when carrying out recognition tasks. But you can't have it both ways. Either the lack of conscious access to the features that we use in making categorizations indicates that we don't make use of features in such tasks; or it indicates the mechanisms that make use of features are not accessible to introspection. But not both.

<sup>15</sup> In fact, Biederman & Shiffrar (1987) report that untrained subjects performed slightly better than chance at chicken sexing, probably because they interpreted a prominent bead as an indication that the chick was male. Many male chicks lack such a bead, so it cannot be used as an accurate cue for maleness, but there is a statistical correlation.

are small and brown, for example). Birds are often timid, and therefore difficult to see. They also move about a lot, so when they are in sight, a quick identification is desirable. And bird-watching is very popular, so lots of people try to develop bird recognition skills.<sup>16</sup>

Now unlike aircraft, tanks, and chicken genitalia, we plausibly do have evolved mechanisms that are specialized for animal-kind recognition. And while we do therefore have some ability to recognise different birds which develops on the basis of casual exposure, as with tank/aircraft identification explicit instruction is required in order to develop the ability to identify a large number of species quickly and accurately, often on the basis of brief and/or distant sightings. Such instruction is widely available in the form of bird books, which point out the features that are diagnostic of particular species.

Experienced bird-watchers develop additional skills. After a great deal of practice, many bird-watchers can identify a bird by the “way it looks”, even when the bird is glimpsed too briefly, or is too far away, to allow individual features to be identified. They even have a name for this brute property, which they call the bird’s “jizz” (see also Tanaka & Curran 2001). This skill takes some time and effort to develop, and is similar to chicken sexing in that birdwatchers report that they perceive the jizz as a gestalt, and cannot say what the features are that make up the whole. Although identifications made on the basis of jizz are reliable, they are generally regarded by bird-watchers as requiring some objective verification.<sup>17</sup>

---

<sup>16</sup> Like many expert skills, bird-watching is not a purely perceptual skill. Expert bird-watchers are able to integrate information from different modalities (appearance and characteristic song, for example), as well as more reflective information such as knowledge of typical habitats and geographical distribution, for example. This suggests that in addition to perceptual detectors, bird-watchers also develop conceptual detectors (see §4.4 above). Here I am concentrating on the perceptual abilities.

<sup>17</sup> One cautionary tale concerns a wader that was spotted in the distance by some bird-watchers, and which had an unusual jizz that none of them recognised. Word spread, and soon a large number of people had gathered to catch sight of the bird, which stubbornly remained too far away to allow a positive identification. Eventually, the bird landed close to the hide, and showed itself to be a very common species that was missing its tail. (Thanks to Gary Allport of BirdLife International for this anecdote, as well as a very interesting conversation on “jizz”.)

### 5.3.3. *Features or gestalt?*

Although non-experts are often amazed by the abilities of experts, there is nothing particularly unusual about these abilities from a psychological point of view. Granted, it takes people a lot of time and effort to learn to reliably identify birds by their jizz, sex chicks or identify aircraft, and these experts are highly accurate, can generally reach a decision quickly and do so subconsciously. In fact, though, we are all constantly making categorizations of this sort: we are highly accurate at recognizing familiar faces, as well as many natural kinds, substances, artefacts, and so on. We do so quickly and subconsciously, and the process is completely inaccessible to introspection. (On what basis do you decide to classify something as a chair or as a tiger, for example? The answer is notoriously difficult to specify.)

If chicken sexers, bird-watchers and plane spotters have developed detection abilities in their particular fields of the same kind that all humans employ to detect many animal kinds, faces, artefacts, and so on, then the various features of these skills that have been reported are exactly what would be expected. When we identify everyday objects and kinds, the nature of the process employed, and the various features that we make use of, are inaccessible to introspection. Our detection abilities are also fast and accurate, and usually a briefly presented or partially occluded object can still be identified. These are precisely the cluster of properties that have been reported for expert detectors in the various domains we have looked at.

The interesting question is then why it is so difficult to acquire these expert skills, and through what process they are acquired. In this regard, it is instructive to consider the processes involved in everyday object categorization (that is, categorization at the 'basic level'; see §4.3.2 above). Certainly, we are not aware of making use of features when we categorize everyday objects—an animal of a particular kind, the face of a friend, the smell of a rose, the sound of a violin. Rather, we have the impression that we categorize 'all-at-once' in more of a gestalt way, rather like an expert bird-watcher identifying a bird by its jizz. But does this mean that everyday categorization is a truly gestalt phenomenon which cannot be broken



down into the recognition of particular features, or is it that such categorization does take place as a result of the recognition of specific features, but that since the process is not accessible to introspection, we are not aware that categorization is feature-based?

Evidence from a number of areas gives strong support to the idea that categorization is based on specific features, but that this is inaccessible to introspection. We will first look at this evidence, before considering what the implications might be for expert categorization, and ultimately for concept acquisition.

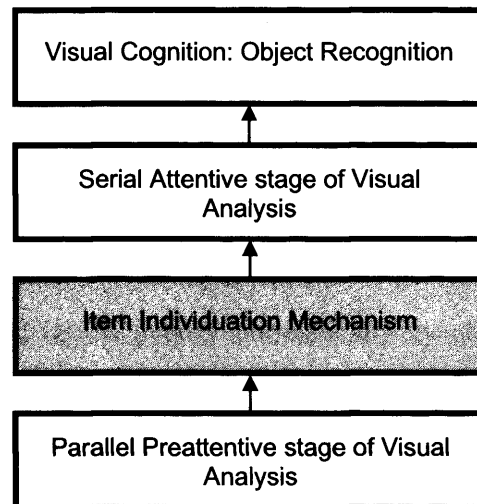
### *Subitizing*

First, consider subitizing, our ability to quickly apprehend small numerosities. This might appear to be a gestalt type of process. It is effortless, extremely fast, and we are not aware of counting the objects in question. Indeed, a number of researchers have in fact proposed that we have neural units that selectively respond to particular numbers of items, up to about 4 (see, for example, the “density theory” proposed by Atkinson et al. 1976).

However, more detailed investigation (Klahr 1973, Trick & Pylyshyn 1993; see also Pylyshyn 2000) has shown that the speed at which we subitize is in fact dependent on the number of items present, with each additional item adding around 60 ms, up to the limit at which we can subitize (around 4 or so; after this, each additional item adds around 200 ms, and errors start to appear). This suggests that there is in fact some sort of “counting” going on, albeit preattentively.

Trick & Pylyshyn (1993) suggest that the evidence on subitizing provides support for a limited-capacity preattentive mechanism for item individuation. They argue that this mechanism must operate at a stage of visual processing which is intermediate between the standardly recognized stages of preattentive visual analysis (which is parallel, hence of unlimited capacity) and attentive visual analysis (which

is serial, hence limited to a single item at a time), as shown in fig. 1 below.<sup>18</sup> This item individuation mechanism has something like an indexical function, in that its role is to track a limited number of individual items (“that thing”) regardless of changes in their location or visual properties.



*Figure 1. Stages of visual processing (adapted from Trick & Pylyshyn 1993: 334).*

So it is reasonably clear that we do not directly apprehend small numerosities. Rather, there is a process of counting, but this is preattentive and not accessible to introspection; only the output of the process is available to consciousness.

#### *Visual word recognition*

Next, consider visual word recognition (Besner & Humphreys 1991; Snowling & Hulme 2005). Reading is a difficult skill to acquire. For accomplished readers, however, the process is fast, accurate, and subconscious—that is, we are not aware while reading of the actual process involved in converting the visual stimuli into meanings. In fact, most people have the impression that they recognize whole words at a glance, rather than having to sound them out.

---

<sup>18</sup> The standard distinction between a parallel preattentive stage and a serial attentive stage in visual analysis was introduced in Treisman’s influential feature-integration theory of attention (Treisman & Gelade 1980).

It is this impression that forms the basis for the “whole-word” (or “whole-language” in its more recent incarnation) approach to the teaching of reading (see Rayner et al. 2002 for a critical discussion). On this approach, children learn by rote how to recognize at a glance a basic vocabulary of words. They then gradually acquire new words through seeing them used in the context of a story; they are encouraged to guess the words they don’t know based on the context rather than by sounding them out. This is in contrast to the other main approach to reading instruction, phonics, which explicitly teaches the connections between letters or letter clusters and phonemes (including exceptions to the standard rules).

A key principle of “whole-language” instruction is that letter-sound correspondences should be discovered automatically by children in the course of exposure to text, and should not be taught explicitly—so much so that the correction of children’s reading errors is actively discouraged. The similarities with the Sargeant system of aircraft recognition discussed in §5.3.2 above are striking (although even E. J. Gibson herself went as far as to say that the whole-word approach to reading was based on Gestalt psychology “applied in a very simple-minded way” which lost all the advantages of an alphabetic language—see E. J. Gibson et al. 1962: 554).

Just as the Sargeant system proved ineffective, experimental evidence suggests that the basic premise of the whole-word approach—that proficient readers recognize whole words at a glance—is false. In a series of experiments, Van Orden and colleagues (Van Orden 1987, 1991; Van Orden et al. 1988; Van Orden & Kloos 2005) began by asking subjects a question, such as “Is it a flower?”. The subject was then presented visually with a word (for example “rose”) and had to indicate whether the word fitted the category. Sometimes, subjects were offered a homophone (either a word or a non-word), such as “rows”. Subjects often mistakenly identified such words as fitting the category, providing evidence that readers routinely convert strings of letters to phonological representations, which they then use to access semantic information for the lexical item. The fact that both word and non-word

homophones were mistakenly identified by subjects as fitting the category provides fairly strong evidence that even accomplished readers access word meanings via phonological representations—that is, they mentally sound out words.

The conclusion that semantic access in reading proceeds via phonology receives further support from a different set of experiments carried out by Lukatela & Turvey (1991, 1993, 1994a, 1994b). In these experiments, subjects were instructed to read aloud as quickly as possible a briefly-presented target word (such as “table”). The target word was preceded by presentation either of a semantically-related word (such as “chair”) or of a non-word homophone of a semantically-related word (such as “chare”). It was found that both semantically-related words and non-word homophones had a priming effect on the speed with which subjects were able to read out the target word. This therefore provides additional evidence that visual word recognition is mediated by phonological representations.<sup>19</sup>

Reviews of the literature comparing the effectiveness of phonics with that of whole-language instruction (such as those reported in Rayner et al. 2002) indicate that while most children will learn the connection between letters and phonemes without explicit instruction, through exposure to text, the explicit teaching of these principles is far more effective. This should not be particularly surprising. Reading is a skill that humans do not have an innate predisposition for. Unlike learning to speak, which develops automatically in almost all humans regardless of general intelligence, learning to read requires explicit instruction and deliberate practice.

---

<sup>19</sup> Other researchers (Coltheart 1980, 2005; Coltheart & Coltheart 1997) have proposed a dual-route theory. The claim here is that there is a processing route directly from orthography to semantics (for skilled readers with well-known words), and that there is also a processing route from orthography to semantics via phonology (for novice readers or with poorly-known words). There are certain experimental results which lend support to this dual-route hypothesis. It doesn't much matter for present purposes how this particular debate turns out, however. The important point is that the 'phonological route only' and the 'dual-route' theories both deny the claim put forward by the whole-word approach, that reading proceeds directly from orthography to semantics in the general case. Both of these theories therefore involve a commitment to a role for phonological features in visual word recognition.

Different teaching methods do produce different results, which is one of the reasons why the debate over the best way to teach reading to children has been so intense.

In visual word recognition, then, we have another case where introspection gives the impression that we recognize a class of stimuli in a gestalt way (that is, by pattern-matching of whole words), whereas in fact recognition is performed on the basis of features (that is, recognition of letters or letter clusters). We are not aware of this fact because the recognition process itself is not accessible to introspection; only the output of the process (that is, the recognized word) is delivered to consciousness.

### *The importance of features*

We can now return to the question that we started with. Is everyday categorization a gestalt phenomenon, or does it involve recognition of specific features? The answer to this question depends, of course, on what we mean by a “feature”. Visual perception proceeds through a series of increasingly abstract processing stages. At the early stages of visual processing, the “simple” neurons of the primary visual cortex identified by Hubel & Wiesel (1962, 1968) respond to a small set of simple stimuli such as lines at particular positions and orientations. At higher-level stages of visual processing, neurons are responsive to much more complex features—including objects such as faces and canonical views of particular living kinds—independent of their specific location in the visual field (Kobatake & Tanaka 1994, Kreiman et al. 2000).

Now, if we consider the simple stimuli detected at early vision to be “features” in the relevant sense, then all perception can be said to be feature-based. But clearly this is not the relevant sense; presumably the thesis that objects are detected on the basis of features is more substantive than that. It would seem rather to be a question about what levels there are intermediate between the simple features of early vision and whole objects. For example, when we recognize a tiger or a chair, is it on the basis of integrating the outputs of all the relevant simple neurons (a gestalt, if you

like), or is it on the basis of identifying component parts (“features” in the present sense)?

The evidence from subitizing and visual word recognition shows that the *process* of recognition itself is not accessible to introspection. We therefore cannot rely on introspection in answering this question, and so will need to look at other forms of evidence.

The above discussion of visual word recognition already suggests that recognition of complex visual stimuli is feature-based rather than gestalt-based. That is, rather than identifying visual words on the basis of complex pattern-matching of the whole word, we make use of letters and letter clusters to recognize the word via its phonological representation in the mental lexicon. It would indeed be surprising if we failed to make use of the valuable information provided by the regular structure of words written in alphabetic scripts (as E. J. Gibson et al. 1962: 554 have pointed out). In particular, in identifying a written word we need to abstract away from large variations in font (as well as upper and lower case) in order to identify the “graphemes” (letter forms) in question (McCandliss et al. 2003), and it would be surprising to say the least if such a process occurred at the level of the whole word, rather than being based on knowledge of the different forms that individual letters can take.

More generally, the importance of features or component parts has been recognized in a number of influential theories of object recognition (Biederman 1987, Hoffman & Singh 1997). Apart from the fact that objects are often made of parts, which can therefore give important clues to object identity (as we saw in the case of written words), recognition of three-dimensional objects may have to be based on identification of parts in many instances. As Hoffman & Singh (1997) point out, objects that are opaque or are partially occluded (that is, nearly all objects encountered in natural settings) present only some of their parts in a single viewing. Animate objects move non-rigidly, presenting a changing visual stimulus. For both of these reasons, parts are indispensable for object identification.

Even in the case of faces, which have often been claimed to be recognized by gestalt, research suggests that recognition depends to a certain extent on the identification of component parts. Again, it would be surprising if our processes of face recognition failed to make use of the regular structure imposed by human physiology. A number of authors have presented evidence that recognition of individual faces relies both on the detailed appearance of component features (parts such as eyes, nose, lips, and so on) and on configural features (the spatial relations among components). See Haig (1984), Sergent (1984), Rhodes (1988), and Carey (1992); see also Rossion et al. (2000).<sup>20</sup> The same is true of ‘Greebles’ (a set of three-dimensional objects representing a family of ‘individuals’ from an artificial kind, distinguishable on the basis of subtle differences in part configuration; see Gauthier & Tarr 1997).

The conclusion we are led to is that the recognition of complex stimuli, even if it appears introspectively to be a gestalt process, is in fact based on the detection of features.

## **5.4. Explaining concept acquisition**

### ***5.4.1. Instinct and learning in categorization***

At the beginning of this chapter we looked at the literature on learning instincts and at how, seen from this perspective, the innate–learned dichotomy breaks down as a characterization of behaviours. We went on to suggest that this dichotomy might also

---

<sup>20</sup> To say that face recognition is feature-based rather than purely gestalt is not to deny that face recognition may be *holistic* (the obligatory processing of all features of an object). A number of studies have suggested that one of the special things about faces as opposed to many other stimuli (animals or houses, say) is that face-recognition is more holistic than recognition of many other stimuli, possibly because configural information is made use of to a greater extent (although some authors, such as Carey & Diamond 1994, claim that there are separate mechanisms for holistic and configural processing). This is highlighted by the fact that own-race faces are recognized more holistically than other-race faces, and upright faces are recognized more holistically than inverted faces (even after considerable training with inverted faces). See Tanaka & Farah (1993), Tanaka & Sengco (1997), Farah et al. (1998), Robbins & McKone (2003) and Tanaka et al. (2004) for discussion.

break down in the case of *ideas*, which could suggest ways to escape the logic of Fodor's (1981) radical concept nativism.

Drawing on the preceding discussion, we will see that the acquisition of lexical concepts, just like behaviours, should be seen as a complex interplay of instinct and learning. As with behaviours, however, it can be useful for expository purposes to give examples where instinct and learning interact to different degrees. In this way, we can begin to understand more clearly what is going on in some of the cases of expert categorization that we looked at above.

For behaviours, we considered three different kinds of situation. We will look at each in turn and explore whether there may be analogues in the case of concepts. The first was behaviour that was innately specified with little possibility for modification through learning, such as stinging behaviour in digger wasps or the avoidance of visual cliffs in infants (E. J. Gibson & Walk 1960). As we noted, innate specification can be vital in some situations (those where there would be no opportunity for learning of an adaptively significant behaviour), but there is a significant cost to the organism in terms of the resulting inflexibility. Similarly, there seem to be some concepts whose contents are innately specified. This would be expected in the case of (a probably limited number of) adaptively significant concepts such as SNAKE. Even in these cases, we noted that there is likely to be some modification through learning of the innately-specified detector, to allow for fine-tuning to the local ecological contingencies (see §4.1.3 above).

The second situation we considered was behaviour that was learned under the guidance of an innately-specified template, such as identification and mapping of nearby nectar sources by bees, or incest avoidance in humans and other primates. Such guidance is needed because there is always an indefinite number of cues to which the organism could attend. Instinctive biases to attend to certain cues are therefore needed in order for anything to be learned at all. We can imagine a similar situation in the case of concepts. The innate component here, rather than being a detector as such (as might be the case for SNAKE), would rather be a set of attention



directors to facilitate acquisition of specific detectors. The great majority of concepts are probably acquired in this way. In the case of animals, for example, an animal-kind template will direct attention to certain cues which are diagnostic of the differences between animal kinds at the basic level (corresponding approximately to the genus), facilitating the acquisition of detectors for these kinds. These attention directors are innately specified and have evolved to reflect those cues that were most stable and adaptively significant in the ancestral environment. This explains how we can acquire concepts for new animal kinds easily, automatically and without specific training. Exposure to kinds (or suitable representations of kinds) is sufficient.

Having an innate template with inbuilt attention directors may not always be enough. In some domains, the diagnostic differences may be too subtle, or too unpredictable to allow for evolved capacities to develop, in which case a different kind of solution is needed. Such is the case with the recognition of individual faces. The face template—itsself activated by certain canonical features of faces, such as a particular configuration of eyes and mouth, in the same way that the animal template is activated by cues to animacy—seems to direct attention to certain first- and second-order features which are particularly diagnostic of the differences between individual faces.<sup>21</sup> This gives some degree of competence in recognizing individual faces from early in life (infants can recognize individual faces from the age of about 4 months, and to a certain extent by 4 days old in the case of their mother's face—Bushnell 1982, Pascalis et al. 1995).

However, the fact that it takes around ten years for adult-level competence in recognizing individual faces to develop (Carey 1992) demonstrates that whatever inbuilt attention directors we have are not sufficient to facilitate individual face

---

<sup>21</sup> Cf. Morton & Johnson (1991) who make a related *ontogenetic* claim that the development of face recognition involves two separate mechanisms: CONSPEC operates in the first 1–2 months and directs the neonate's attention to face-like stimuli, after which CONLERN, which is responsible for acquiring knowledge about individual faces, gradually takes over.

recognition with the same degree of ease as for animal kind recognition, say.<sup>22</sup> This is in spite of the fact that we plausibly have an innate module specifically adapted to the task.<sup>23</sup> The prolonged acquisition period suggests that something more is going on than mere feature detection on the basis of inbuilt attention directors. Plausibly, in this case some of the cues themselves need to be learned. That is, while attention may be directed to certain component features (eyes, nose, and so on), and to certain higher-order (configural) features, the child still needs to refine and add to these higher-order cues so that they are diagnostic of the differences between the individual faces encountered in the local environment. A number of lines of evidence converge on this conclusion. First, it has been shown that it is not the matching of a face to its stored representation that children have difficulty with, but the actual encoding of faces—that is, the extraction of features (Carey 1992). Second, it has been shown that children are less affected by face-inversion than adults. Since inversion affects configural features rather than local features (Bartlett & Searcy 1993, Searcy & Bartlett 1996), this suggests that it is the encoding of configural information that children are poor at, implying that this is a key skill that is fine-tuned in the development of face expertise (Diamond & Carey 1986, Carey 1992, Carey & Diamond 1994, Gauthier & Tarr 1997, Pascalis et al. 2001).<sup>24</sup> Lastly, prosopagnosics may have intact recognition of face components, and to the extent that they are able to recognize individual faces, it is by slow and laborious matching of component features (Bliem & Danek 1999); they have been shown to perform

---

<sup>22</sup> The difficulty of faces compared to animal kinds is of course partly due to the different levels of categorization involved (the individual level and basic level, respectively), a point to which we will return below.

<sup>23</sup> Many studies have shown that face processing is special (see Farah et al. 1998, Nelson 2001), although there remains some debate over whether the relevant domain is faces or the broader class of those stimuli requiring fine visual discrimination (Gauthier et al. 2000, Gauthier & Nelson 2001). Evidence presented in Farah et al. (2000) concerning a case of prosopagnosia in a 16-year-old subject who sustained brain damage as a newborn strongly suggests that a dedicated neural substrate for face processing may be specified in the genome.

<sup>24</sup> But note that children of all ages (even in infancy) are still affected by face inversion to a certain extent, so it cannot be the case that children initially make use of component features exclusively. See Carey & Diamond (1994).

*better* at matching inverted faces than at matching normal faces, perhaps because upright faces automatically trigger their (impaired) face module (Farah et al. 1995). Again, these data suggest that it is configural information that is critical to face recognition, and impaired in prosopagnosia.

So, in many cases innate attention directors will be sufficient for the rapid and automatic acquisition of appropriate detectors, and hence for the acquisition of the corresponding concepts. Evidence from individual face recognition, however, shows that in some cases (when the required distinctions are too subtle, or subject to too much contingent variation) acquiring detectors requires learning which cues are diagnostic, or at the very least fine-tuning the cues to which we are sensitive. This has relevance for the acquisition of expert perceptual skills, which we will discuss in more detail below.

The third situation we considered was behaviour that was learned without significant innate guidance, which we noted was probably extremely rare other than in humans (“cultural” behaviours in some apes being one possible exception). In the case of concepts, this situation plausibly arises for a number of expert domains where we have no inbuilt bias towards the relevant cues, meaning that we have to be taught the cues or learn them through extensive self-study. The difficulty, though, is how to distinguish this third situation from skills like individual face recognition, which also take several years to perfect. The question we need to answer is whether expert skills like chicken sexing or aircraft recognition take years to perfect because they are inherently difficult (because the relevant perceptual distinctions are so subtle), or whether the problem is that we lack the necessary inbuilt guidance towards the diagnostic cues.

There are a number of relevant considerations here. First, people don’t automatically acquire expert abilities from mere exposure to stimuli. This is different from the situation with individual face recognition, since this ability develops automatically and without explicit training. (Face recognition does, however, take several years to perfect, suggesting that even though the relevant cues are not inbuilt,

a predisposition to acquire these cues is.) Now, it could be argued that the same is true in the case of expert skills, and that the fact that these skills do not develop spontaneously is merely because most people do not naturally encounter the relevant stimuli (aircraft, chicken genitalia, and so on) on a regular enough basis, so there is no opportunity for the relevant abilities to develop. On this view, the only difference between individual face recognition and, say, aircraft recognition, is that most of us are not exposed to enough aircraft to develop the necessary skills. There may indeed be something to this claim, but it is unlikely to be the whole story. Recall that this claim was more-or-less the position of the Gibsonian 'ecological' psychologists, who thought that, given sufficient exposure to the right sorts of stimuli, expert perceptual skills could be acquired automatically in the same way that we acquire our everyday categorization skills (see §5.3 above). It is almost certain that such an approach would not have worked with chicken sexing: humans have been farming poultry for millennia, but it was only in the 1920s on the basis of detailed morphological study that a perceptual basis for such a discrimination was discovered. It is therefore very unlikely that mere repeated exposure to chick genitalia would have been sufficient. And even in the case of aircraft recognition (not to mention reading), where the Gibsonian approach had a certain limited amount of success, it proved to be far less effective than explicit instruction. Further, a number of studies have shown that mere exposure to stimuli is not sufficient for the development of perceptual expertise (Ericsson et al. 1993).

Indeed, it does seem plausible that the difficulty that untrained subjects have with aircraft recognition is not due to any inherent difficulty with the task, but is due rather to a failure to attend to the diagnostic cues. This is clear from the fact that aircraft are readily identifiable from their silhouettes by experts and by automatic target recognition systems (Jaggi et al. 1999), indicating that global shape cues are sufficient for recognition. Experienced bird-watchers and fish experts are also able to recognize objects in their domain of expertise at the subordinate level on the basis of silhouettes, again demonstrating that such stimuli are not inherently difficult to

identify (Johnson 1992, Johnson & Mervis 1997). We are all able to easily identify many everyday objects on the basis of their silhouettes (Hoffman & Singh 1997), but this is obviously not true of recognizing individual faces, where configural information is vital (something which makes automatic face recognition far more of a technical challenge—see Hancock et al. 2000). Consider also the case of cars. Most urban dwellers are exposed to many different models of car on a daily basis without developing the recognition skills of an automobile enthusiast, even after many years of such exposure. Rather than being a reflection of the difficulty of the task, it seems that this is due to the fact that most people just don't automatically attend to the right cues—that is, we don't have a “vehicle” template which directs our attention to the diagnostic cues so that we can automatically develop car-type detectors.

A number of researchers have looked at the question of which cues we attend to under which circumstances. Tanaka & Taylor (1991) have shown that acquiring expertise in a particular domain may have the effect of changing the “basic level” for that domain. That is, experts categorize objects in their domain of expertise at a more subordinate level than non-experts. In support of this, Archambault et al. (1999) have shown that the perceptual features that people employ for representing objects depends on the level at which they categorize those objects. Thus, people will pay attention to different features of an object when they categorize it at the basic level than when they categorize that same object at the subordinate level (cf. Johnson & Mervis 1997). Schyns & Murphy (1994) have argued that the need to make finer perceptual discriminations to distinguish previously conflated categories prompts the learning of new object cues.

There is an apparent contradiction in the literature on expertise. On the one hand, many authors underline the fact that expert skills involve the ability to make fine perceptual judgements, often on the basis of configural cues rather than component features, and take many years to acquire (Tanaka & Taylor 1991, Carey 1992, Ericsson & Lehmann 1996). On the other hand, a number of authors have investigated ‘expert’ abilities in the laboratory by training novices in particular tasks.

For example, Biederman & Shiffrar (1987) trained chick sexing 'experts' by giving novices a short instructional note explaining how to differentiate (pictures of) male and female chick genitalia. Similarly, Gauthier and colleagues (Gauthier & Tarr 1997, Gauthier et al. 1998) investigated the neural basis for 'Greeble' recognition in 'experts' they had trained for a period of less than 10 hours each. And Tanaka et al. (2005) trained subjects to be 'experts' in the recognition of various bird species (all from a single family) over several consecutive days. But surely, if these kinds of expert abilities normally take up to a decade to acquire, even when considerable training is given (as in the case of chicken sexing, for example), how can experimental subjects be trained in such skills over the course of just a few hours?

A plausible explanation is that acquiring expert recognition skills can involve developing several different kinds of abilities, some of which are more straightforward than others. Consider again the case of aircraft recognition. As noted above, the perceptual basis for distinguishing aircraft does not seem to be inherently that difficult. Thus, in order to achieve a reasonable level of expertise it will be sufficient to have our attention directed to the relevant diagnostic cues (through explicit training). The same should be true for recognizing cars, birds, fish and so on at the subordinate level.<sup>25</sup> However, to obtain further improvements in accuracy or speed it will be necessary for subjects to learn subtle new perceptual cues, possibly based on higher-order configural information. These are not easily taught, and may require modifications in our visual perception over long periods of exposure to relevant stimuli. This is the question to which we now turn.

---

<sup>25</sup> Note that the subjects trained in chick sexing by Biederman & Shiffrar (1987) could not be considered to have acquired expert skills, for a number of reasons. First, they were able to take as long as they liked to complete the sexing task. Second, they were working with written instructions and test sheet, and so could work out the answer using conscious reflection and explicit reference to the instructions. Third, they were informed that the test photographs consisted of an equal number of male and female chicks. Fourth, their accuracy was nowhere near the 99 per cent plus required of commercial chicken sexers. It is precisely in order to internalise the relevant perceptual distinctions and develop the required speed and accuracy that expert chicken sexers must develop their skills over several years.

#### **5.4.2. *Perceptual expertise and perceptual learning***

There are numerous studies showing that under certain conditions changes can be induced in a perceptual system which enable improved discriminations to be made (for a review, see Goldstone 1998). A number of mechanisms have been identified. *Attentional weighting* adapts perception to particular environments or tasks by directing selective attention to diagnostic perceptual features (and hence also reducing attention to irrelevant features). This or a closely related process is responsible for the acquisition of phonological categories (Werker & Tees 1984, Maye et al. 2002), and similar effects for visual categorization have been demonstrated in the laboratory (Goldstone 1994). *Stimulus imprinting* is a process whereby specialized detectors develop for particular repeated stimuli or features. For example, Schyns & Rodet (1997) have found that arbitrary curved parts recurring in a set of artificial stimuli are more likely to be used as a basis for categorization if they have been important in earlier tasks (see also Schyns & Murphy 1994). *Differentiation* separates previously indistinguishable stimuli or features, allowing them to be perceptually distinguished. This can occur for simple features associated with early perceptual processing such as line segments at particular orientations (Ahissar & Hochstein 1993, 1996), or higher-level features such as species of bird (Tanaka et al. 2005). Lastly, *unitization* works in the opposite direction to differentiation, by ‘chunking’ stimulus features together so that they become a single perceptual unit (again, this can occur with simple or higher-level features). This is the process underlying the encoding of configural features for (upright) faces and other stimuli (Diamond & Carey 1986, Young et al. 1987, Carey & Diamond 1994, Gauthier & Tarr 1997, Gauthier et al. 1998).

What is the nature of the perceptual learning underlying expert perceptual abilities? It is likely that all four of the mechanisms discussed above are involved. In the case of expert abilities such as aircraft recognition or bird-watching, attentional weighting is clearly important, as we noted above. Once we have learned to attend to the diagnostic cues relevant to these domains, in many circumstances the task of

discrimination is not particularly difficult. This is why even a person with little previous experience can find field guides of use in distinguishing bird species (or other living kinds), since these guides explicitly point out diagnostic features. However, the development of expertise does not stop at this point (cf. Williams et al. 1998). Experts are not only able to direct attention to the diagnostic cues, they are also able to make discriminations on the basis of more subtle stimulus features, such as configural features, which involves more than just the direction of selective attention. This allows experts to further increase their speed and accuracy, including in more difficult situations (such as when the stimulus is briefly presented, viewed from a distance, or degraded in some other way). It is at this point that processes of differentiation and unitization are important.<sup>26</sup>

In order to understand how these processes of differentiation and unitization work in the case of perceptual expertise, it is first necessary to know whether the development of expertise involves modifications to the simple features of early perception or to higher-level features. Most of the research on such perceptual learning has involved laboratory investigations of simple perceptual tasks such as visual search. For example, Ahissar & Hochstein (1993, 1996) have shown that practice can improve performance in “pop-out” search tasks (when an odd item appears to pop-out of an array of distractors). Since pop-out is generally taken to be mediated by the parallel preattentive stage of visual processing, this implies that practice in the task produces modifications to low-level visual features. However, as Tanaka et al. (2005) point out, there are a number of considerations which suggest that this cannot be the process underlying perceptual expertise. First, modifications at this low level have been shown to be specific to the stimuli used in training, so that practice in one task will not bring about improvements in other closely related

---

<sup>26</sup> The only mechanism I have not discussed is stimulus imprinting. This would be the mechanism responsible for triggering innate concepts (see §1.3.3 above), and could possibly have a role in the acquisition of detectors for certain classes of object that are identified on the basis not of constituent parts but of viewpoint-dependent representations of the whole object. I will not speculate further here, but see Palmeri (1997).



tasks (Ahissar & Hochstein 1993). By contrast, perceptual expertise generalises to new contexts (for example, expertise with birds facilitates discrimination of previously unencountered bird species—Tanaka et al. 2005). Second, low-level visual learning can be acquired very quickly (Ahissar & Hochstein 1997). However, as we have seen, perceptual expertise develops over a period of years. Finally, it is not altogether clear whether, or how, improvements in detecting simple low-level features would translate into skills at recognizing the complex objects of expert domains.

At the same time, there are tantalizing indications that low-level perceptual learning and the development of perceptual expertise are related processes. The first such indication relates to the role of selective attention. As we have seen in §5.3.2 above, the development of perceptual expertise in domains such as chicken sexing and aircraft recognition requires that selective attention be directed to the task. Ahissar & Hochstein (1993) showed that in the case of perceptual learning also, improvements occurred only for those aspects of low-level visual stimuli to which subjects selectively attended, and not for unattended aspects of the same stimuli. This latter result is surprising, however, because the early visual processes apparently being modified are preattentive, so it is not clear what the crucial role of attention could be here. A second indication that perceptual learning and perceptual expertise might be related comes from the extent to which these skills are generalizable. While it is broadly true, as noted above, that perceptual learning does not generalize to the same extent that perceptual expertise does, the situation is by no means straightforward. While perceptual expertise may extend to new categories within the domain of expertise, it is clear that expertise does not transfer across domains (expertise in bird-watching does not transfer to dogs, and vice-versa—Tanaka & Taylor 1991). And even within the domain of expertise there are limits. For example, expertise with dogs has been shown to be specific to familiar breeds (Diamond & Carey 1986), and our expertise at individual face recognition does not extend to faces of unfamiliar races (Chance et al. 1982). As for perceptual learning,

it has been found that it is possible to manipulate the generalizability of the learning by adjusting the difficulty of the learning context. Ahissar & Hochstein (1997) demonstrated that increasing the difficulty of the learning situation (for example, by reducing the stimulus presentation time) led to far more specificity, whereas in easy situations the learning was far more generalized.<sup>27</sup> These and other findings led Ahissar and Hochstein to develop a new theory for visual perceptual learning, ‘reverse hierarchy theory’ (Ahissar & Hochstein 1997, 2004; Ahissar 1999).

According to reverse hierarchy theory, visual perceptual learning is a top-down attention-driven process for selecting diagnostic cues.<sup>28</sup> It begins at high-level vision, but if no suitable cues are found, it then works its way down the visual system to progressively lower levels. This process can also induce modifications to the visual system at each level, proceeding in reverse hierarchical order—essentially, a process of ‘tuning’ the relevant neuronal populations to the properties of the task stimuli.<sup>29</sup> This reverse-hierarchical order makes ecological sense, since high-level neurons represent more abstract features or whole objects having higher ecological validity (but at the cost of a reduction in fine detail). The theory makes several predictions about the perceptual learning process, and can explain a number of puzzling results:

---

<sup>27</sup> Note, by the way, that this finding explains the problems with the Renshaw system for aircraft recognition discussed in §5.3.2 above. Recall that this system employed very brief tachistoscopic presentation as its training method. The findings of Ahissar and Hochstein suggest that since this is a difficult training condition, the resultant learning will not generalize. And indeed, the reason that the system was abandoned was that while trainees greatly improved their performance under task conditions, it was found that this did not translate into significant improvements under natural conditions. Cf. also Ashworth & Dror (2000), who found that learning of distinctive aircraft showed greater generalization to novel orientations than aircraft that were similar in appearance.

<sup>28</sup> Ahissar and Hochstein do not exclude the possibility that there may also be bottom-up processes of modification to the visual system. This may be the source of perceptual learning in the absence of attention (as reported, for example, in Watanabe et al. 2001).

<sup>29</sup> It used to be thought, following the pioneering experiments of Hubel and Wiesel, that low-level visual areas lost their plasticity early. Recent research now suggests that the primary visual area of adult mammals in fact retains considerable plasticity. See Gilbert & Wiesel (1992).

*Generalizability.* This theory explains the finding that learning in easy conditions generalizes more than learning in hard cases. Easy cases are those for which the cues are easily accessed. According to reverse hierarchy theory, it is the high-level cues which are accessed first. These are based on the fairly coarse but more generalized discriminations made by high-level neuronal populations, which explains why learning in these contexts is more generalizable. When the task is more difficult, it is necessary to access lower-level cues, which are based on the much finer but more specific discriminations made by the neurons of low-level vision. Thus, learning under these conditions is less able to be generalized.

*Learning proceeds from easy to hard cases.* The above considerations furthermore suggest that learning will be most effective if it starts with easier cases and then gradually proceeds to more difficult cases. This is consistent with experimental findings by Ahissar & Hochstein (1997).

*Speed of learning.* The fact that learning proceeds from easy to hard cases suggests that learning easy tasks, which involves modifications at higher levels, will be faster. Learning which requires modifications at lower levels will be much slower (or in some cases impossible if it has not been preceded by learning of a similar but easier task). This has been confirmed experimentally (see Ahissar & Hochstein 1997, 2004; Karni & Sagi 1993).

*The role of attention.* In reverse hierarchy theory, selective attention is the process whereby the cues relevant to the discrimination task at hand are identified. This explains why perceptual learning requires selective attention. Consistent with this view, it has been found that attention is required in order for learned improvements in “pop-out” detection to occur. However, the speed with which learning was achieved in these

tasks (together with a certain degree of generalizability) are inconsistent with the normal assumption that pop-out occurs at early vision. Accordingly, Ahissar and Hochstein posit that pop-out is a relatively high-level phenomenon. This conclusion is independently suggested by a range of other findings—for example, the fact that low-level neurons are sensitive to precise position information whereas pop-out is not, and the fact that such neurons make fine orientation or colour discriminations that do not pop out (see Hochstein & Ahissar 2002, Wolfe 2003).

Reverse hierarchy theory also allows an integrated view of perceptual learning and perceptual expertise. In particular, the extensive deliberate practice involved in the development of perceptual expertise in a particular domain will result, according to Ahissar & Hochstein (2004), in the following four phases of learning:

*Phase 1.* The discriminations of novices are governed by the high-level cues that are employed for quotidian recognition tasks (and which have high ecological validity in most contexts).

*Phase 2.* Following relevant training and some exposure to a particular expert domain, there is general improvement in making the relevant perceptual discriminations. This is achieved by increasing the accessibility of diagnostic cues—for example, by attentional weighting and via mechanisms to guide attention to diagnostic cues at lower levels, which are identified via a guided search down the activated pathways to lower processing levels. (These attention-guidance mechanisms must obviously be flexible, since different cues are diagnostic in different contexts; we will discuss this further in the following section.)

*Phase 3.* After a great deal of training and deliberate practice in a given domain, neuronal populations at different levels start to become tuned to the particular domain-specific demands, improving performance in

difficult cases of discrimination. This is achieved by changing the integration properties of higher-level neurons (increasing the weighting of inputs relevant to the task, and decreasing the weighting of less relevant inputs). These changes occur initially at the stage of high-level perception, but then gradually proceed to lower levels, creating a chain of modifications to the particular neuronal populations at various levels which are selective for the relevant discrimination tasks. Thus, the cues at each level become increasingly diagnostic for domain-specific discriminations.

*Phase 4.* After a number of years of consistent practice, the process of modification to neuronal populations at successively lower levels leads to the development of more highly diagnostic cues at higher and higher levels. This is because changes at low levels also affect all higher levels which receive input from them. Eventually, therefore, experts are able to rely once again on high-level cues, even for difficult discriminations. In essence, their high-level stage of perception is now biased towards the expert domain.<sup>30</sup>

This proposal has the potential to account for many of the properties of perceptual expertise that we have discussed earlier. As we have seen, perceptual expertise takes several years to develop, which is of course explained by the fact that the process outlined above is one of slow, gradual development in the course of extended practice.

It is also clear that perceptual expertise does not develop spontaneously or automatically. It normally requires training in the form of explicit instruction to highlight important features of the stimuli. In the absence of training, a great deal of

---

<sup>30</sup> It is a common finding that the development of expertise comes at a cost to flexibility—for example, the Stroop effect (Stroop 1935), and the effects reported in Gauthier & Tarr (1997) and Young et al. (1987).

motivated self-study is needed. This is the case for the ‘self-taught’ experts that we encountered earlier, such as those pioneering chicken sexers of the 1920s, or the early aviation enthusiasts (famously, sushi chefs are also supposed to acquire their skills in this way; see de Waal 2001). We can now see why this should be the case. In domains such as chicken sexing or aircraft recognition for which we have no inbuilt attention directors, training (or extensive self-study) is necessary in order to draw attention to the diagnostic cues. Without this crucial step, we would be unable to initiate the process of attentional weighting, for example, and hence would be unable to achieve the second phase of expertise outlined above. This is a precondition for development of further expertise according to the reverse hierarchy theory.

In the case of individual face recognition, the situation is slightly different. Here we do have inbuilt attention directors to the diagnostic cues (including component and configural features). The presence of inbuilt attention directors in the case of faces explains why some recognition ability is present from an early age. It also explains why explicit instruction is not required for the further development of this skill: evolution has jump-started the process by endowing us with the abilities required for phase-2 expertise. A lengthy learning period is then still required to automatically tune the system to the particular properties of the local face population.

Specialized natural kind discrimination (birds, fish, plants, trees, and so on) perhaps falls somewhere between individual face recognition and chicken sexing. At the basic level (approximately, the genus), we have inbuilt attention directors to the diagnostic cues, and discrimination is fast and automatic. (Although modern urbanites may have had so little experience with some of these categories that they are only able to make discriminations readily at the life-form level—‘bird’, ‘tree’ or ‘bush’, for example; see Johnson & Mervis 1997, Atran et al. 2004.) When we need to make discriminations at the species level, these inbuilt mechanisms are usually no longer sufficient. We may need explicit instruction (say, from field guides) in order

to direct our attention to the cues that are diagnostic at this level, or at least a great deal of experience (for example, Berlin 1992: §§5.5–5.6 reports that Aguaruna hunters could discriminate animals at the species level, but non-hunters could not). Once we are aware of the diagnostic cues (that is, once we have achieved phase-2 competence), we can then develop greater levels of expertise given sufficient practice.

In addition to explicit instruction, one of the hallmarks of perceptual expertise is that it requires a great deal of deliberate practice (see §5.3.2 above). In particular, studies have found that the levels of expertise attained after years of experience alone are much lower than in the case of experts who have performed regular deliberate practice (Ericsson et al. 1993). This can be explained from the perspective of reverse hierarchy theory on the basis of the fundamental role of selective attention. Deliberate practice ensures that selective attention is directed at the relevant aspects of the task, whereas experience alone does not. The direction of selective attention is necessary in order for the various perceptual modifications underlying the development of expertise to be made. This is confirmed by experimental results. For example, Tanaka et al. (2005) found that subjects who learned to classify a set of birds at the family level developed much poorer discrimination abilities than subjects who learned to classify the same birds at the species level.

A further feature of perceptual expertise is that the recognition process becomes intuitive after a certain level of expertise is reached, in the sense that experts are not aware of the cues that they are utilizing. This is notoriously the case for chicken sexers, but the same is true of experts at aircraft recognition (Allan 1958) and in other domains (see Myers 2002). It is also clearly true of everyday recognition at the basic level—as noted previously, we are not generally aware of the basis on which we recognize objects. As Herbert Simon (1992: 155) has put it: “intuition is nothing more and nothing less than recognition”. Although we do not have conscious access to the cues that we use in the course of recognition, the need for explicit instruction

in order to achieve phase-2 competence means that our attention will be consciously drawn to the diagnostic cues in these cases. For this reason, we will be aware of the cues that we are using. However, once the learning and neuronal modification process takes off, the cues that we make use of are no longer those that we were trained to attend to; in particular, once considerable expertise is achieved, recognition will be once again based on high-level cues (Ahissar & Hochstein 2004: 462) and will have the fast and intuitive feel of everyday object recognition (as, for example, when accomplished bird-watchers can identify a bird by its 'jizz'). It is these factors which give rise to the intuitive nature of perceptual discriminations once a certain level of expertise has been reached. (It is interesting to note that in the case of the Sargeant system of aircraft recognition training discussed in §5.3.2 above, which did not make use of explicit training, it is reported that trainees were able to recognize aircraft without being conscious of using component features.) The fact that experts can identify subordinate-level objects as quickly as basic-level objects (Tanaka & Taylor 1991) is a reflection of the fact that they have reached the final phase of expertise and are able to use high-level cues to distinguish objects in their domain of expertise, just as they would for basic-level objects.

Lastly, consider the 'holistic' nature of perceptual expertise. This term is given various interpretations in the perceptual expertise literature; for present purposes I shall take it to mean the obligatory processing of all features of an object, due to greater reliance on configural cues.<sup>31</sup> Since inversion has been shown to disrupt holistic/configural processing, the extent to which recognition in particular domains is holistic can be tested with inverted stimuli (Diamond & Carey 1986; Tanaka & Farah 1993). It has been found that individual face recognition becomes increasingly holistic as a function of expertise (Farah et al. 1998), and the same is true of recognition abilities in other domains of perceptual expertise (Gauthier et al. 2000,

---

<sup>31</sup> It may turn out that we need to make a distinction between 'holistic' and 'configural', as argued by Carey & Diamond (1994), but it is not crucial to take a stand on this issue here. Cf. footnote 20 (p. 196) above.



2003). A reliance on configural cues is driven by the need to make individual-level discriminations of homogeneous stimuli (individual faces, individual dogs, and so on). Given that configural information is based on the spatial relations between the parts of an object, it follows that configural cues must be based on high-level perceptual representations. Therefore holistic processing, which makes use of configural information, is a reflection of considerable perceptual expertise in homogeneous domains. In the case of non-homogeneous stimuli—such as objects at the basic level—configural cues will be much less necessary, which explains why everyday objects are not processed holistically, as evinced by the fact that we are far less sensitive to inversion of such objects (Diamond & Carey 1986).<sup>32</sup>

In the next section, we will take a step back from the detail and look at the implications of perceptual learning for the acquisition of concepts. But first, a brief philosophical point. The preceding considerations about perceptual learning may tempt one towards relativism. After all, it's only a short step from the claim that experts have reconfigured aspects of their visual perception to the claim that experts actually perceive the world in different ways (to see just how short, take a look at Goldstone 1994). In a sense, this is trivially true. A bird-watcher may see a lesser spotted woodpecker where a novice just sees a small black bird, and an adult may recognize a familiar face that a child does not. But so what? The stronger claim, that the novice and the expert or the child and the adult have fundamentally different conscious experience of the world, requires a much stronger argument. And while I don't intend to provide detailed arguments here, there seems to me no reason to

---

<sup>32</sup> There is a potential objection to the picture I have presented. If holistic processing is a reflection of a high level of expertise which normally takes years to acquire (except when nature gives us a head start, as in the case of faces), how can we explain the findings of Gauthier and colleagues that subjects who had been trained for a few hours in Greeble recognition showed sensitivity to configural changes (Gauthier & Tarr 1997, Gauthier et al. 1998)? One possibility is that the sensitivity shown by subjects was to changes in individual features rather than global configuration, particularly since the orientation of certain parts was a cue to Greeble 'gender', and hence individual Greeble identity. Thus, inverting the Greeble would disrupt detection of this (part-based rather than configural) cue. (Broader questions are raised by the neuroanatomical data that Gauthier and colleagues present suggesting similarities between Greeble processing and face processing.)

assume that neurological changes which improve the detection of configural cues to face identity, say, need have any fundamental effect on how faces actually *appear* to us. Perceptual learning occurs only in certain highly constrained ways, and the effect (or even the very purpose) of these constraints may be to prevent changes which would fundamentally alter how the world appears; if this were not the case, perceptual learning could be very dangerous for the organism.<sup>33</sup>

#### ***5.4.3. The acquisition of perceptual detectors***

It is often the case that investigating extreme examples can inform thinking about more regular cases. The development of perceptual expertise represents an extreme form of concept acquisition, but I have spent some time on the details because I believe that it can help to shed light on concept acquisition more generally.

It is important to underline that the acquisition of regular, everyday object concepts does not involve perceptual learning. Our perceptual systems have been tuned by evolutionary processes precisely for the task of distinguishing between objects at the basic level. In a sense, we are all perceptual experts in the domain of everyday objects, and there is little that we can do to improve further. Experimental studies have confirmed this. For example, it has been shown that humans (as well as monkeys) can detect the presence of an animal in a novel natural scene in as little as 150 ms (Thorpe et al. 1996). This performance cannot be improved by several weeks of intensive training—trained subjects performed just as well on completely novel scenes as on the scenes that had been used during training (Fabre-Thorpe et al. 2001).

So what is the mechanism underlying the acquisition of everyday object concepts? As discussed earlier, in a number of cases we plausibly have hard-wired detectors, meaning that the corresponding concepts are innate. For the rest, acquiring the concept involves the acquisition of a corresponding detector. This is a process

---

<sup>33</sup> This is not to say that such pathological changes cannot be *induced*—by trauma, say, or during development through the sorts of extreme environmental manipulations that Hubel and Wiesel inflicted on their cats.

which operates under strong innate guidance. The innate contribution is the conceptual template, which includes among other things a set of domain-specific inferential procedures (including meaning postulates), and attention directors to cues that are diagnostic for the perceptual discrimination of the entities in question.

To take a concrete example, consider what might take place when we encounter a novel kind of animal. Certain cues to animacy (the presence of eyes, say) will cause our animal-kind template to be triggered (cf. Atran 1998). This template includes inbuilt attention directors to certain diagnostic cues. Experience of individual members of this animal kind will enable the particular values for these cues to be specified. The template thus prompts the learning of a detector for this particular animal kind.

Perceptual learning is a process of transcending the innate constraints imposed by a template. It can involve changing the relative weighting of the cues for those cases where the default weighting is non-optimal (this is phase 2 of the Ahissar and Hochstein account of perceptual expertise acquisition sketched above). This could be the case where a particular class of objects triggers a template that is not well-suited to the task.<sup>34</sup> Perceptual learning can also involve learning additional cues when those specified in the template are not optimal. For example, the individual-face template plausibly has slots not only for the values of particular cues, but also for the precise form of some of the cues themselves, to allow for race-specific configural information to be utilised.<sup>35</sup>

---

<sup>34</sup> For example, suppose that monkey faces trigger our human-face template. Then developing expertise with monkey faces may require changing the weighting of some cues we use for human faces.

<sup>35</sup> A third possibility, which I will not discuss further, would be to change the default partitioning of the set of possible values for a particular cue (see the discussion of 'binning' in Berretty et al. 1999). For example, imagine a binary-valued cue for tail length. This would involve partitioning the set of all tail-length values into two (that is, 'long' versus 'short'). We could imagine developing a three-way partition (that is, 'long', 'medium', and 'short'), or higher partitions, without otherwise changing the cue or its weighting.

Innate templates make good sense, of course, as does the adaptability to changing environments and novel perceptual tasks that perceptual learning provides. There are an infinite number of cues that the organism could potentially attend to in any given situation, so some guidance is obviously required. But there is also a need for flexibility: “adapt or die”. The psychological literature on change blindness gives an obvious demonstration of the fact that we only attend to a very limited range of the available information (cf. Archambault et al. 1999).

Cognitive efficiency involves not only deciding which particular information to attend to in a given situation, but also how to process that information most effectively. Just as we can be overloaded by too much information, we can also be overloaded by excessive processing demands. To be efficient, perceptual detection therefore needs not only attention directors, but also effective algorithms with which to process that information. In what follows I want to consider one particular proposal for categorization which has been put forward within the “fast and frugal heuristics” framework (Gigerenzer et al. 1999). While I am not arguing in favour of this particular model, the functional considerations which motivate such models can give insight into fundamental aspects of the problem, and this can complement the insights from empirical studies that we looked at in the previous section. (I do also feel that the particular approach advocated is a promising one.)

As we have seen, perceptual detection at the basic level is *fast*—which is just as well, since the survival of the organism could be at stake. This suggests that the categorization decision should be made on basis of the smallest possible number of cues. A suitable stopping rule will further facilitate rapid categorization decisions: the organism should stop processing cues as soon as it is able to reach a decision, rather than always considering the full range of prescribed cues. Berretty et al. (1999) propose a categorization procedure which would fulfil these requirements, and which they term “categorization by elimination” (cf. Tversky 1972).

The procedure works as follows. Cues are accessed sequentially in a pre-determined order, and each cue eliminates candidates from the set of possible

categories for an object (initially, this is the set of all categories). When only one category remains, the procedure stops, and the object is assigned to this category. In the case where all cues are exhausted, and more than one possible category remains, a random assignment is made (although we could imagine other possibilities in this situation: the procedure could offer no categorization, or could return a classification at a superordinate level, say).

With a categorization procedure such as this, cue order has a significant impact on how effective the procedure is (particularly as regards speed; accuracy is less affected by changes in cue order). It is therefore important for organisms to be able to consider cues in the optimal order. It would thus be expected that innate templates would impose an ordering on the cues that they specify, and that this ordering would be relatively fixed. (Note that in a model of this kind, cue weighting establishes the cue order, whereas in a Bayesian model, for example, the numerical weighting itself is used in computing the decision).

The expectation that templates would impose a fixed cue order is consistent with ethological findings. For example, we noted earlier that honey-bees apply a fixed set of cues in a fixed order when identifying flowers—odour, colour and shape, in that order. Not only is the cue order fixed, but honey-bees are also unable to add to the predetermined set of cues. Thus, although the direction of light polarization is a cue which is diagnostic of flower identity, and bees are highly sensitive to this property in other circumstances (such as for navigation), they are unable to use it to identify flowers. We are not honey-bees, of course, and the possibility for perceptual learning allows us to adapt to novel situations. But this adaptability is itself subject to innate constraints, which further research on perceptual learning can help to uncover. There are certain things, such as a veridical perception of the Müller-Lyer illusion, that we just can't learn.

The assumption that this categorization procedure initially accesses the set of all categories may appear to run counter to the idea that we have domain-specific reasoning abilities. But such a model of categorization is not necessarily

incompatible with domain specificity.<sup>36</sup> We can see the initial cues (animacy, say, or intentionality) as being used to select the appropriate domain, after which the template for that particular domain is triggered, and is the basis for determining the subsequent cues that are used.

Thus far, we have considered how perceptual detectors might be acquired—both in the standard case where it is only the values of a predetermined set of cues that need to be supplied, and the special case where the cue order or the cues themselves need to be acquired via perceptual learning. However, it is clear that the acquisition of perceptual detectors can't be the *whole* story about concept acquisition. For one thing, as we saw in chapters 2 and 3, it explains only one aspect of content (referential content) and has nothing to say about the other aspect of content (logical content). For another thing, it explains only perceptual detection, and has nothing to say about conceptual detection (a distinction discussed in chapter 4). Before closing, then, a few remarks on each of these points.

As to the first point, we need to consider how meaning postulates (or procedures more generally) might be acquired. Given the potential cognitive risks associated with acquiring new intuitive inference rules, we would expect that this possibility is at least as heavily constrained as perceptual learning. Indeed, one option would be to claim simply that we cannot acquire new meaning postulates. But limiting us in this way to only those meaning postulates that are specified in innate templates would probably be too strict. It seems that we at least want to leave open the possibility that after a great deal of practice, we could come to internalise a reflective rule of inference (think of the logician who after years of applying *modus tollens* in a reflective way, eventually develops the ability to reason intuitively with it; or consider the scientist who develops a theory and who, after years of grappling in a reflective way with the principles and implications of the theory eventually comes to have an intuitive grasp of it).

---

<sup>36</sup> See Samuels & Stich (2004) for discussion of heuristics and modularity.

Now, it may be that meaning postulates can be extracted from experience automatically, by a kind of analogue to perceptual learning. There are certain tentative indications from the literature on implicit learning that this might be the case. Implicit learning is a process by which covariant relationships and possibly even relatively complex abstract rules can be learned, without there being any conscious awareness on the part of the learner that this has taken place, or any conscious access to what has been learned (see Lewicki et al. 1992; Cleeremans et al. 2000). Indeed, there do seem to be considerable parallels between implicit learning and perceptual learning. As with perceptual learning, it has been shown that implicit learning requires selective attention to relevant, predictive stimulus information (Jiang & Chun 2001), and learning takes place in a gradual way. If it turns out that rules of inference can be learned implicitly, then the constraints on the acquisition of meaning postulates could be somewhat similar to the constraints on perceptual learning. If so, this would place considerable limits on the kinds of meaning postulates that could be acquired—they would be limited, that is, to expressing statistical or covariance regularities. This would probably not allow for cases such as the logician or the scientist that we considered above. Another possibility, then, might be that meaning postulates or procedures that we initially learn to apply reflectively could over time and practice become intuitive. This question deserves further investigation. (For some related thoughts, see Peacocke 2004: 93 f.)

As to the second point, we need to consider the acquisition of non-perceptual concepts. For example, artefact concepts rely on the conceptual (rather than perceptual) detection of ‘original intended function’. Now, in some cases this may be possible on the basis of perceptual cues, as discussed in chapter 4, and so we may have perceptual detectors associated with some artefact concepts. But in many other cases, the detection of function will have to rely on inferences outside the scope of our perceptual processes. Concepts which are fully abstract will rely solely on such inferences. This conceptual detection will be intuitive (as when we identify someone on the basis of a combination of face and voice cues), and may well be based on

templates (/heuristics), in the same way as perceptual detection. Consider, for example, the concept LIE (see Coleman & Kay 1981). When people are provided with a range of particular examples, there is broad intersubjective agreement about what counts as a lie. However, most people are unable to verbalise the basis on which they decide that something is a lie.<sup>37</sup> This suggests that we may have some intuitive—but clearly not perceptual—detector for lies, the basis for which is not (readily) accessible to introspection. In other cases, concepts may be permanently reflective, due to the fact that it will never be possible to grasp them intuitively (religious concepts, say, as discussed in chapter 4).

Much more needs to be said on both of these points, but it is getting late, so I will leave these questions for future research.

### **5.5. Conclusions: What are concepts and how are they acquired?**

Over the last five chapters, we have followed a reasonably long and occasionally winding course. In this final section, I want to give an overview of the various arguments presented in the thesis, and an integrated view of what the implications are for philosophical accounts of conceptual content and for psychological accounts of concept acquisition. I'll take things in roughly that order.

The thesis has discussed a number of challenges to Fodor's theory of informational atomism. Recall that informational atomism is the conjunction of two theses:

*informational semantics*, according to which content is constituted exhaustively by nomological mind–world relations; and

---

<sup>37</sup> Thus, most people will say that a lie is any false statement. A moment's reflection shows this cannot be the case: a false statement need not be a lie (if the speaker believes it to be true, say, or if there is no intent to deceive), and a lie need not be false (the liar may unknowingly be saying something true).



*conceptual atomism*, according to which (lexical) concepts have no internal structure.<sup>38</sup>

One of the main arguments of this thesis has been that we need to supplement informational semantics by allowing content-constitutive rules of inference (meaning postulates). The need for such inference rules is clear, and was lucidly pointed out by Carroll (1895)—without inference rules we have no account of inference; for representational theories of mind more generally, we have no account of psychological processes without computational rules. I don't take this to be contentious. The only contentious issue is whether such meaning postulates should be seen as content constitutive.

Since rules of inference have commonly been seen as the only way to account for the content of logical terms, one possibility would be to allow content-constitutive meaning postulates for the logical vocabulary, but to deny that meaning postulates are constitutive of the content of the non-logical vocabulary. This was the position that Fodor defended until recently. The difficulty is that it requires drawing a principled distinction between the logical and non-logical vocabularies, and as we have seen the prospects for this do not look promising. This means that an uncomfortable choice must be made. Either one is forced to completely abandon the idea that meaning postulates are content constitutive, and find some other—and completely novel—account for how logical terms get their content. This is the

---

<sup>38</sup> Fodor sometimes presents atomism in stronger terms, as the thesis that “satisfying the metaphysically necessary conditions for having one concept *never* requires satisfying the metaphysically necessary conditions for having any other concept” (1998a: 13–14, original emphasis). This stronger thesis is what you get when you accept *both* conceptual atomism and informational semantics (which of course Fodor does). To see why, consider that a conceptual atomist can allow content-constitutive meaning postulates without undermining the thesis that (lexical) concepts have no internal structure. But since meaning postulates must often be formulated in terms of more than one concept, it would follow that possessing a particular concept often *did* require possession of some other concept. However, informational semantics precisely *denies* that content is constituted by anything other than mind–world links, so an atomist who also adopts informational semantics will indeed deny that the metaphysics of concept possession is in any way dependent on what other concepts one has. See Fodor (1998a: Appendix 5A) for some discussion.

position taken by Fodor (2004a, 2004b). The other option is to accept that meaning postulates are constitutive of the content of both logical and non-logical terms. This is the position I have argued for in this thesis.

Although I have suggested that such a claim has a good deal of inherent plausibility, we immediately face a well-known problem. If certain inferences are content constitutive (the meaning postulates), then we need to find a principled way of telling these inferences from the rest. The obvious way to do so is by employing some notion of analyticity. However, this possibility is called into serious doubt by Quine's well-known and widely-accepted arguments against any principled analytic/synthetic distinction.

Proponents of analyticity required it to do some heavy-duty philosophical work, in particular in serving as a justification for a priori knowledge. Our current aims are far more modest: we need some notion which will support the distinction between meaning postulates and other inferences. I have argued in chapter 2 that such a distinction can be made on the basis of the psychological division noted earlier between representation and computation. On this basis I developed a notion of 'psychosemantic analyticity', and showed that since this is underwritten by empirical psychological criteria, it is immune from Quine's arguments against analyticity. (However, there is nothing which ensures that our mental representations must be veridical, hence no guarantee that a meaning postulate will be valid; this notion of psychosemantic analyticity can certainly not underwrite a priori knowledge.) The psychological characterization of meaning postulates that I have proposed suggests some possible constraints on which inferences could be in fact governed by meaning postulates. First, any meaning postulate, since it is a rule of inference, will be regarded by the individual as intuitively valid. Second, following Sperber & Wilson (1995) I have suggested that meaning postulates will be limited to elimination rules (that is, there will be no introduction rules).

Suppose that you were convinced up to this point. If so, you might nevertheless be troubled by the following concern. An inferential-role account for the content of

the Boolean connectives, say, is standardly given in terms of the canonical introduction and elimination rules for those connectives. How are we to account for the meaning of these terms if we also deny that there are any introduction rules?<sup>39</sup> Based on detailed and psychologically-motivated proposals for how inference rules serve to fix content, I demonstrated in chapter 3 that it is possible to account for the content of logical connectives exclusively in terms of elimination rules.<sup>40</sup> Furthermore, I showed on the basis of these proposals that no particular set of inference rules is required in order to fix the content of a logical connective (in fact, none would be needed at all if we could directly represent truth tables, and reason on this basis).

At first blush, this might be taken to show that Fodor is right that the possession of logical concepts is not a matter of the inferences that such concepts enter into, but is rather just a matter of having a concept with the right content. I have taken it to show something different. That is, I argued that *if* a particular meaning postulate is attached to a logical concept then it contributes to fixing the content of that concept. However, since different sets of meaning postulates can converge on the same logical content, no identification of meaning postulates with possession conditions can be made. Possession of a particular logical concept does not entail possession of a fixed set of meaning postulates. I further argued that since no principled distinction can be made between the logical vocabulary and the rest then, by extension, meaning postulates in general are content constitutive. The difference between logical and non-logical terms is that, while meaning postulates exhaust the content of the former, they do not exhaust the content of the latter.

---

<sup>39</sup> You might also be concerned about how to account for logical inference without introduction rules—the very reason for proposing meaning postulates in the first place. As we have seen in §3.5, this concern can be taken care of by postulating certain non-standard—but empirically supported—elimination rules.

<sup>40</sup> We considered, and rejected, the possibility that meaning postulates could be seen as some kind of mechanism for sustaining a content-constitutive link between concepts and abstract logical properties.

Notice that advocating a loosening of *informational semantics* by allowing content-constitutive meaning postulates in no way undermines *conceptual atomism*. Quite the opposite—meaning postulates can in fact provide additional support to an atomistic account of lexical concepts, by addressing a number of problems that would otherwise necessitate a move away from conceptual atomism.

There are a number of problematic cases for informational atomism that we have looked at in this thesis. We have seen that accounting for these cases often requires that a choice be made between maintaining informational semantics at the expense of conceptual atomism, or vice-versa. Fodor regularly takes the former option, at the cost of accepting what he hopes is a limited amount of atomism for lexical concepts. This across-the-board commitment to informational semantics costs Fodor dear, though, and I'm not sure in the end it's worth the price.<sup>41</sup>

For example, his solution to the case of necessarily coreferential concepts such as TRIANGLE/TRILATERAL is to maintain an informational semantic account, but allow that the modes of presentation are complex. He takes the same approach for context-restricted concepts such as ADDLED, maintaining an informational semantic account on which ADDLED and SPOILED (say) are synonymous, but allowing that having the concept EGG is a possession condition for having ADDLED. And a similar approach is adopted to deal with cross-linguistic cases where a single (unambiguous) word in one language translates, depending on context, as one of two (or more) words in a second language. Finally, we have the case of concepts for nomologically impossible entities. By sticking to an informational semantics for these concepts, Fodor is forced to abandon atomism and claim that such concepts are complex (that is, phrasal). It is debatable, though, whether the range of cases considered above can really be treated

---

<sup>41</sup> In any case, such an across-the-board commitment cannot be maintained. There are certain concepts, as we saw in chapter 2, where Fodor accepts that it is not possible to give a purely informational semantic account. This was the case for non-synonymous concepts expressing necessarily coinstantiated properties, as with Quine's case of RABBIT and UNDETACHED PROPER RABBIT PART (and perhaps such pairs as BUY and SELL). These present fewer problems for an account that can employ meaning postulates.

as rare exceptions to an otherwise strictly atomic view of lexical concepts. I have argued that supplementing informational semantics with meaning postulates can offer natural solutions to a number of these problematic cases.

It is important to be clear that in arguing for meaning postulates as mentally-represented content-constitutive rules of inference, I am not claiming that *all* mental inference is governed by meaning postulates—in particular, reflective thinking often requires applying rules that we do not have an intuitive grasp of (modus tollens, say). Also, I am not claiming that *all* concepts have meaning postulates attached—again, reflective thinking often involves concepts that are not deployed in intuitive reasoning (theoretical or religious concepts, say). In chapter 4 we looked more closely at the distinction between intuitive and reflective beliefs, and the corresponding distinction between intuitive and reflective concepts, as developed by Dan Sperber. Intuitive concepts include concepts for objects which can be identified by our perceptual processes, as well as additional (generally more abstract) concepts required for the representation of meaning postulates. Reflective concepts occur in metarepresentational contexts and are thereby insulated from our intuitive thought processes.

The thesis went on to consider what the view of conceptual content set out above could tell us about concept acquisition. Notoriously, Fodor has in the past endorsed the radically nativist position that all lexical concepts are innate. The informational atomism that he has developed, based on an externalist approach to content, allows him to move away from this radical position. Rather than requiring that all lexical concepts be innate, informational atomism requires only that certain mechanisms are innate—those mechanisms, that is, which are needed to ensure that we reliably get locked to the property of *Xness* in the presence of stereotypical Xs.

In chapters 4 and 5, I looked at one way of cashing out this notion of ‘reliably getting locked to the property of *Xness*’. One obvious mechanism (but not the only one) that would ensure this is an inbuilt disposition to acquire detectors for the objects that we encounter. The acquisition of detectors could also explain how a

reflective concept (acquired, say, via communication) could become intuitive. I investigated in some detail how the acquisition of perceptual detectors might work.<sup>42</sup> In doing so, I drew on recent work in ethology on ‘learning instincts’, and recent work on the neurological basis for perceptual learning. What emerged was an understanding of how one kind of mechanism of the type suggested by Fodor might work. We saw that perceptual detectors involve a complex interplay between innate constraints and environmental input. Thus, in some limited cases (SNAKE, say), it may be that we do possess innate detectors, and hence have the corresponding concepts innately. In most cases (animal kind concepts at the basic level, say), the cues that are diagnostic for kind discrimination are inbuilt and only the values for these cues must be learned. In other more exceptional cases (such as occurs in the development of perceptual expertise), acquiring detectors involves learning both the cues and the cue values. In support of Fodor’s position, then, the acquisition of perceptual detectors requires innate mechanisms (such as those responsible for directing attention to cues), but does not require that concepts themselves be innate. (A question that we raised, but did not discuss in detail, was whether meaning postulates should all be seen as innate, and if not what process might underlie their acquisition.)

\* \* \*

What, then, are the prospects for informational atomism as a theory of concepts? Jerry Fodor once commented that inferential role semantics “*having once got a reasonable story about ‘and’, took it for granted that ‘tree’ would submit to much the same treatment. But it doesn’t.*”<sup>43</sup> It is reasonable in my view to make a symmetrical comment concerning informational atomism. Fodor, after all, has a

---

<sup>42</sup> I also noted that some detectors would operate at the conceptual rather than perceptual level. This would be true, for example, of modality-independent detectors that operated by integrating the outputs of modality-specific perceptual detectors. Some process of this kind must underlie our ability to take both face and voice cues into consideration when identifying a familiar person, say.

<sup>43</sup> Personal communication, 2000.

pretty good story to tell about 'tree', but informational atomism requires that this story extend also to 'and'. In my view it doesn't. The reasons why it doesn't, and what form a possible solution could take, have been the primary preoccupation of this thesis.

## Bibliography

- Achinstein, P. 1974. The identity conditions of properties. *American Philosophical Quarterly*, 11: 257–275.
- Ahissar, M. 1999. Perceptual learning. *Current Directions in Psychological Science*, 8: 124–128.
- Ahissar, M. & S. Hochstein 1993. Attentional control of early perceptual learning. *Proceedings of the National Academy of Sciences of the United States of America*, 90: 5718–5722.
- Ahissar, M. & S. Hochstein 1996. Learning pop-out detection: specificities to stimulus characteristics. *Vision Research*, 36: 3487–3500.
- Ahissar, M. & S. Hochstein 1997. Task difficulty and the specificity of perceptual learning. *Nature*, 387: 401–406.
- Ahissar, M. & S. Hochstein 2004. The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, 8: 457–464.
- Allan, M. D. 1958. Learning perceptual skills: the Sargeant system of recognition training. *Occupational Psychology*, 32: 245–252.
- Anderson, J. R. & C. Lebiere 2003. The Newell Test for a theory of cognition. *Behavioral and Brain Sciences*, 26: 587–601.
- Archambault, A., C. O'Donnell & P. G. Schyns 1999. Blind to object changes: when learning the same object at different levels of categorization modifies its perception. *Psychological Science*, 10: 249–255.
- Armstrong, D. 1997. *A World of States of Affairs*. Cambridge: Cambridge University Press.
- Ashworth, A. R. S. & I. E. Dror 2000. Object identification as a function of discriminability and learning presentations: the effect of stimulus similarity and canonical frame alignment on aircraft identification. *Journal of Experimental Psychology: Applied*, 6: 148–157.



- Atkinson, J., F. Campbell & M. Francis 1976. The magic number 4 + 0: a new look at visual numerosity judgements. *Perception*, 5: 327–334.
- Atran, S. 1998. Folk biology and the anthropology of science: cognitive universals and cultural particulars. *Behavioral and Brain Sciences*, 21: 547–609.
- Atran, S. 2002. Modular and cultural factors in biological understanding: an experimental approach to the cognitive basis of science. In P. Carruthers, S. Stich & M. Siegal (eds) *The Cognitive Basis of Science*, 41–72. Cambridge: Cambridge University Press.
- Atran, S., D. Medin & N. Ross 2004. Evolution and devolution of knowledge: a tale of two biologies. *Journal of the Royal Anthropological Institute*, 10: 395–420.
- Aydede, M. 2000. On the type/token relation of mental representations. *Facta Philosophica*, 2: 23–49.
- Baayen, R. H., R. Piepenbrock & L. Gulikers 1995. *The CELEX lexical database* [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.
- Bach, K. 2000. Review of Fodor (1998a). *The Philosophical Review*, 109: 627–632.
- Baillargeon, R. 2002. The acquisition of physical knowledge in infancy: a summary in eight lessons. In P. K. Smith & C. H. Hart (eds) *Blackwell Handbook of Childhood Cognitive Development*, 47–83. Oxford: Blackwell.
- Baker, L. R. 1991. Has content been naturalized? In Loewer & Rey (eds) 1991, 17–32.
- Barrouillet, P., N. Grosset & J.-F. Lecas 2000. Conditional reasoning by mental models: chronometric and developmental evidence. *Cognition*, 75: 237–266.
- Barrouillet, P. & J.-F. Lecas 1998. How can mental models theory account for content effects in conditional reasoning? A developmental perspective. *Cognition*, 67: 209–253.
- Bartlett, J. C. & J. H. Searcy 1993. Inversion and configuration of faces. *Cognitive Psychology*, 25: 281–316.
- Barwise, J. & J. Perry 1981. *Situations and Attitudes*. Cambridge, MA: MIT Press.

- Berdoy, M., J. P. Webster & D. W. Macdonald 2000. Fatal attraction in rats infected with *Toxoplasma gondii*. *Proceedings of the Royal Society of London (Series B: Biological Sciences)*, 267: 1591–1594.
- Berlin, B. 1992. *Ethnobiological Classification: Principles of Categorization of Plants and Animals in Traditional Societies*. Princeton, NJ: Princeton University Press.
- Berlin, B., D. E. Breedlove & P. H. Raven 1973. General principles of classification and nomenclature in folk biology. *American Anthropologist*, 75: 214–242.
- Berretty, P. M., P. M. Todd & L. Martignon 1999. Categorization by elimination: using few cues to choose. In Gigerenzer et al. 1999, 235–254.
- Besner, D. & G. Humphreys (eds) 1991. *Basic Processes in Reading: Visual Word Recognition*. Hillsdale, NJ: Erlbaum.
- Biederman, I. 1987. Recognition-by-components: a theory of human image understanding. *Psychological Review*, 94: 115–147.
- Biederman, I. & M. M. Shiffrar 1987. Sexing day-old chicks: a case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13: 640–645.
- Blakemore, D. 2000. Indicators and procedures: ‘nevertheless’ and ‘but’. *Journal of Linguistics*, 36: 463–486.
- Blanchette, P. A. 2001. Logical consequence. In L. Goble (ed.) *The Blackwell Guide to Philosophical Logic*, 115–135. Oxford: Blackwell.
- Bliem, H. R. & A. Danek 1999. Direct evidence for a consistent dissociation between structural facial discrimination and facial individuation in prosopagnosia. *Brain & Cognition*, 40: 48–52.
- Block, N. 1986. Advertisement for a semantics for psychology. In P. French, T. Uehling & H. Wettstein (eds.) *Midwest Studies in Philosophy, Volume 10: Studies in the Philosophy of Mind*, 615–678. Minneapolis: University of Minnesota Press.

- Block, N. 1993. Holism, hyper-analyticity and hyper-compositionality. *Mind and Language*, 8: 1–26.
- Bloom, P. 1996. Intention, history, and artifact concepts. *Cognition*, 60: 1–29.
- Bloom, P. 1998. Theories of artifact categorisation. *Cognition*, 66: 87–93.
- Bloom, P. 2000. *How Children Learn the Meanings of Words*. Cambridge, MA: MIT Press.
- Boghossian, P. A. 1991. Naturalizing content. In Loewer & Rey (eds) 1991, 65–86.
- Boghossian, P. A. 1993. Does an inferential role semantics rest upon a mistake? *Mind and Language*, 8: 27–40.
- Boghossian, P. A. 1994. Inferential role semantics and the analytic/synthetic distinction. *Philosophical Studies*, 73: 109–122.
- Boghossian, P. A. 1997. Analyticity. In B. Hale & C. Wright (eds) *A Companion to the Philosophy of Language*, 331–368. Oxford: Blackwell.
- Booth, A. E. & S. Waxman 2002. Object names and object functions serve as cues to categories for infants. *Developmental Psychology*, 38: 948–957.
- Braine, M. D. S. & D. P. O'Brien 1991. A theory of *if*: A lexical entry, reasoning program, and pragmatic principles. *Psychological Review*, 98: 182–203.
- Brandom, R. B. 1998. Insights and blindspots of reliabilism. *The Monist*, 81: 371–392.
- Brook, A. & R. J. Stainton 1997. Fodor's new theory of content and computation. *Mind and Language*, 12: 459–474.
- Brown, R. 1958. How shall a thing be called? *Psychological Review*, 65: 14–21.
- Burge, T. 1979. Individualism and the mental. In P. A. French, T. E. Uehling & H. K. Wettstein (eds) *Studies in Metaphysics*. Midwest Studies in Philosophy, vol. IV, 73–121. Minneapolis: University of Minnesota Press.
- Burton, A. M., V. Bruce & R. A. Johnston 1990. Understanding face recognition with an interactive activation model. *British Journal of Psychology*, 81: 361–380.

- Bushnell, I. W. R. 1982. Discrimination of faces by young infants. *Journal of Experimental Child Psychology*, 33: 298–308.
- Cain, M. J. 2002. *Fodor: Language, Mind and Philosophy*. Cambridge, UK: Polity.
- Caramazza, A. & J. R. Shelton 1998. Domain-specific knowledge systems in the brain: the animate-inanimate distinction. *Journal of Cognitive Neuroscience*, 10: 1–34.
- Carey, S. 1992. Becoming a face expert. *Philosophical Transactions of the Royal Society of London (Series B: Biological Sciences)*, 335: 95–102.
- Carey, S. & R. Diamond 1994. Are faces perceived as configurations more by adults than by children? *Visual Cognition*, 1: 253–274.
- Carnap, R. 1952. Meaning postulates. *Philosophical Studies*, 3: 65–73.
- Carnap, R. 1956. *Meaning and Necessity*, enlarged edition. Chicago: University of Chicago Press.
- Carroll, L. 1895. What the tortoise said to Achilles. *Mind*, 4: 278–280.
- Carruthers, P., S. Laurence & S. Stich (eds) 2005. *The Innate Mind: Structure and Content*. Oxford: Oxford University Press.
- Carston, R. 1998. Informativeness, relevance and scalar implicature. In R. Carston & S. Uchida (eds) *Relevance Theory: Applications and Implications*, 179–236. Amsterdam: John Benjamins.
- Carston, R. 2002. *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell.
- Chance, J. E., A. L. Turner & A. G. Goldstein 1982. Development of differential recognition for own- and other-race faces. *Journal of Psychology*, 112: 29–37.
- Chierchia, G. & S. McConnell-Ginet 2000. *Meaning and Grammar*, 2<sup>nd</sup> edition. Cambridge, MA: MIT Press.
- Clark, H. H. 1969a. Linguistic processes in deductive reasoning. *Psychological Review*, 76: 387–404.
- Clark, H. H. 1969b. The influence of language in solving three-term series problems. *Journal of Experimental Psychology*, 82: 205–215.

- Clark, H. H. 1971. More about “adjectives, comparatives, and syllogisms”: a reply to Huttenlocher and Higgins. *Psychological Review*, 78: 505–514.
- Cleeremans, A., A. Destrebecqz & M. Boyer 2000. Implicit learning: news from the front. *Trends in Cognitive Sciences*, 2: 406–416.
- Coleman, L. & P. Kay 1981. Prototype semantics: the English word *lie*. *Language*, 57: 26–44.
- Coltheart, M. 1980. Reading, phonological recoding and deep dyslexia. In M. Coltheart, K. E. Patterson & J. Marshall (eds) *Deep Dyslexia*, 197–226. London: Routledge & Kegan Paul.
- Coltheart, M. 2005. Modelling reading: the dual-route approach. In Snowling & Hulme (eds) 2005, 6–23.
- Coltheart, M. & V. Coltheart 1997. Reading comprehension is not exclusively reliant upon phonological representation. *Cognitive Neuropsychology*, 14: 167–175.
- Cormack, A. 1998. *Definitions: Implications for Syntax, Semantics, and the Language of Thought*. New York: Garland.
- Cosmides, L. & J. Tooby 2000a. Consider the source: the evolution of adaptations for decoupling and metarepresentation. In Sperber (ed.) 2000, 53–115.
- Cosmides, L. & J. Tooby 2000b. Evolutionary psychology and the emotions. In M. Lewis & J. M. Haviland-Jones (eds) *Handbook of Emotions*, 2<sup>nd</sup> edition, 91–115. New York: Guilford.
- Cosmides, L., J. Tooby & R. Kurzban 2003. Perceptions of race. *Trends in Cognitive Sciences*, 7: 173–179.
- Cowie, F. 1999. *What's Within?: Nativism Reconsidered*. Oxford: Oxford University Press.
- Crain, S., A. Gualmini & P. Pietroski 2005. Brass tacks in linguistic theory: innate grammatical principles. In Carruthers et al. (eds) 2005, 175–197.
- Cresswell, M. J. 1978. Semantic competence. In M. Guenther-Reutter & F. Guenther (eds) *Meaning and Translation*, 9–28. London: Duckworth.

- Currie, G. 2000. Imagination, delusion and hallucinations. *Mind and Language*, 15: 168–183.
- de Almeida, R. G. 1999. What do category-specific semantic deficits tell us about the representation of lexical concepts? *Brain and Language*, 68: 241–248.
- De Renzi, E. & F. Lucchelli 1994. Are semantic systems separately represented in the brain?: The case of living category impairment. *Cortex*, 30: 3–25.
- de Waal, F. B. M. 1999a. The end of nature versus nurture. *Scientific American*, 281.6 (December 1999): 56–61.
- de Waal, F. B. M. 1999b. Cultural primatology comes of age. *Nature*, 399: 635–636.
- de Waal, F. B. M. 2001. *The Ape and the Sushi Master*. New York: Basic Books.
- Dehaene, S. 1997. *The Number Sense*. New York: Oxford University Press.
- Dennett, D. C. 1975. Brain writing and mind reading. In K. Gunderson (ed.) *Minnesota Studies in the Philosophy of Science, Volume VII: Language, Mind, and Knowledge*, 403–416. Minneapolis: University of Minnesota Press.
- Dennett, D. C. 1990. The interpretation of texts, people and other artifacts. *Philosophy and Phenomenological Research*, 50 (Supplement): 177–194.
- Diamond, R. & S. Carey 1986. Why faces are and are not special: an effect of expertise. *Journal of Experimental Psychology: General*, 115: 107–117.
- Dowty, D. 1979. *Word Meaning and Montague Grammar*. Dordrecht: Reidel.
- Dretske, F. 1977. Laws of nature. *Philosophy of Science*, 44: 248–268.
- Dretske, F. 1981. *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.
- Dretske, F. 1983. Précis of *Knowledge and the Flow of Information*. *Behavioral and Brain Sciences*, 6: 55–90. (Includes commentaries and author's response)
- Dretske, F. 1986. Misrepresentation. In R. Bogdan (ed.) *Belief*, 17–36. Oxford: Oxford University Press.
- Dretske, F. 1989. Reasons and causes. *Philosophical Perspectives*, 3: 1–15.
- Drewery, A. 2005. Essentialism and the necessity of the laws of nature. *Synthese*, 144: 381–396.

- Dreyfus, H. L. 1990. Socratic and platonic sources of cognitivism. In J.-C. Smith (ed.) *Historical Foundations of Cognitive Science*, 1–17. Dordrecht: Kluwer.
- Ekman, P., M. O’Sullivan & M. G. Frank 1999. A few can catch a liar. *Psychological Science*, 10: 263–266.
- Ellis, H. D., D. M. Jones & N. Mosdell 1997. Intra- and inter-modal repetition priming of familiar faces and voices. *British Journal of Psychology*, 88: 143–156.
- Epstein, S. 1994. Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, 49: 709–724.
- Epstein, S., A. Lipson, C. Holstein & E. Huh 1992. Irrational reactions to negative outcomes: evidence for two conceptual systems. *Journal of Personality and Social Psychology*, 62: 328–339.
- Ericsson, K. A., R. T. Krampe & C. Tesch-Römer 1993. The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100: 363–406.
- Ericsson, K. A. & A. C. Lehmann 1996. Expert and exceptional performance: evidence of maximal adaptation to task constraints. *Annual Review of Psychology*, 47: 273–305.
- Evans, J. St. B. T., S. J. Handley & D. E. Over 2003. Conditionals and conditional probability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29: 321–335.
- Evans, J. St. B. T., S. E. Newstead & R. M. J. Byrne 1993. *Human Reasoning: The Psychology of Deduction*. Hove, UK: Lawrence Erlbaum.
- Evans, J. St. B. T. & D. E. Over 2004. *If*. Oxford: Oxford University Press.
- Everett, A. 2005. Recent defences of descriptivism. *Mind & Language*, 20: 103–139.
- Fabre-Thorpe, M., A. Delorme, C. Marlot & S. Thorpe 2001. A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of Cognitive Neuroscience*, 13: 171–180.
- Farah, M. J. 2004. *Visual Agnosia*, 2<sup>nd</sup> edition. Cambridge, MA: MIT Press.

- Farah, M. J., C. Rabinowitz, G. E. Quinn & G. T. Liu 2000. Early commitment of neural substrates for face recognition. *Cognitive Neuropsychology*, 17: 117–123.
- Farah, M. J., K. D. Wilson, H. M. Drain & J. R. Tanaka 1995. The inverted face inversion effect in prosopagnosia: evidence for mandatory, face-specific perceptual mechanisms. *Vision Research*, 35: 2089–2093.
- Farah, M. J., K. D. Wilson, M. Drain & J. N. Tanaka 1998. What is “special” about face perception? *Psychological Review*, 105: 482–498.
- Field, H. 1977. Logic, meaning and conceptual role. *Journal of Philosophy*, 74: 379–409.
- Fine, K. 1982. First-order modal theories. III: Facts. *Synthese*, 53: 43–122.
- Fodor, J. A. 1970. Three reasons for not deriving “kill” from “cause to die”. *Linguistic Inquiry*, 1: 429–438.
- Fodor, J. A. 1975. *The Language of Thought*. New York: Crowell.
- Fodor, J. A. 1978. Propositional attitudes. In *Representations: Philosophical Essays on the Foundations of Cognitive Science*, 177–203. Cambridge, MA: MIT Press.
- Fodor, J. A. 1980. Fixation of belief and concept acquisition. In M. Piattelli-Palmarini (ed.) *Language and Learning: The Debate between Jean Piaget and Noam Chomsky*, 143–149. London: Routledge & Kegan Paul.
- Fodor, J. A. 1981. The present status of the innateness controversy. In *Representations: Philosophical Essays on the Foundations of Cognitive Science*, 257–316. Cambridge, MA: MIT Press.
- Fodor, J. A. 1982. Cognitive science and the Twin-Earth problem. *Notre Dame Journal of Formal Logic*, 23: 98–118.
- Fodor, J. A. 1983. *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.
- Fodor, J. A. 1984a. Semantics, Wisconsin style. *Synthese*, 59: 231–250.



- Fodor, J. A. 1984b/1990. Psychosemantics or: Where do truth conditions come from? Ms., MIT. (Published in W. G. Lycan (ed.) *Mind and Cognition*, 312–337. Oxford: Blackwell, 1990.)
- Fodor, J. A. 1984c. Observation reconsidered. *Philosophy of Science*, 51: 23–43.
- Fodor, J. A. 1987. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press.
- Fodor, J. A. 1990a. A theory of content, I: The problem. In *A Theory of Content and Other Essays*, 51–87. Cambridge, MA: MIT Press.
- Fodor, J. A. 1990b. A theory of content, II: The theory. In *A Theory of Content and Other Essays*, 89–136. Cambridge, MA: MIT Press.
- Fodor, J. A. 1990c. Information and representation. In P. Hanson (ed.) *Information, Language, and Cognition*, 175–190. Vancouver: University of British Columbia Press.
- Fodor, J. A. 1991. A modal argument for narrow content. *Journal of Philosophy*, 88: 5–26.
- Fodor, J. A. 1994. *The Elm and the Expert: Mentalese and its Semantics*. Cambridge, MA: MIT Press.
- Fodor, J. A. 1998a. *Concepts: Where Cognitive Science Went Wrong*. Oxford: Oxford University Press.
- Fodor, J. A. 1998b. There are no recognitional concepts—not even RED. *Philosophical Issues*, 9: 1–14.
- Fodor, J. A. 2000a. Replies to critics. *Mind and Language*, 15: 350–374.
- Fodor, J. A. 2000b. *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*. Cambridge, MA: MIT Press.
- Fodor, J. A. 2001a. Doing without What's Within: Fiona Cowie's critique of nativism. *Mind*, 107: 99–148.
- Fodor, J. A. 2001b. Language, thought and compositionality. *Mind and Language*, 16: 1–15.
- Fodor, J. A. 2003a. *Hume Variations*. Oxford: Oxford University Press.

- Fodor, J. A. 2003b. Crossed wires: the link between what's in the genes and what's in the mind. Review of Ridley (2003). *Times Literary Supplement*, 16 May 2003: 3–4.
- Fodor, J. A. 2004a. Having concepts: a brief refutation of the twentieth century. *Mind and Language*, 19: 29–47.
- Fodor, J. A. 2004b. Reply to commentators. *Mind and Language*, 19: 99–112.
- Fodor, J. A. 2004c. Revenge of the given. Ms., Rutgers.
- Fodor, J. A., M. F. Garrett, E. C. T. Walker & C. H. Parkes 1980. Against definitions. *Cognition*, 8: 263–367.
- Fodor, J. A. & E. Lepore 1991. Why meaning (probably) isn't conceptual role. *Mind and Language*, 6: 328–343.
- Fodor, J. A. & E. Lepore 1992. *Holism: A Shopper's Guide*. Oxford: Blackwell.
- Fodor, J. A. & E. Lepore 2002. *The Compositionality Papers*. Oxford: Oxford University Press.
- Fodor, J. A. & E. Lepore 2005. Out of context. *Proceedings and Addresses of the American Philosophical Association*, 78: 3–20.
- Fodor, J. A. & Z. W. Pylyshyn 1981. How direct is visual perception?: Some reflections on Gibson's "Ecological Approach". *Cognition*, 9: 139–196.
- Fodor, J. D. 1977. *Semantics: Theories of Meaning in Generative Grammar*. New York: Crowell.
- Fodor, J. D., J. A. Fodor & M. F. Garrett 1975. The psychological unreality of semantic representations. *Linguistic Inquiry*, 4: 515–531.
- Galef, B. G. 1988. Imitation in animals. In T. R. Sentall & B. G. Galef (eds) *Social Learning: Psychological and Biological Perspectives*, 3–28. Hillsdale, NJ: Erlbaum.
- Garcia, J. & R. A. Koelling 1966. The relation of cue to consequence in avoidance learning. *Psychonomic Science*, 4: 123–124.
- Gates, G. 1996. The price of information. *Synthese*, 107: 325–347.

- Gauthier, I., T. Curran, K. M. Curby & D. Collins 2003. Perceptual interference supports a non-modular account of face processing. *Nature Neuroscience*, 6: 428–432.
- Gauthier, I. & C. A. Nelson 2001. The development of face expertise. *Current Opinion in Neurobiology*, 11: 219–224.
- Gauthier, I., P. Skudlarski, J. C. Gore & A. W. Anderson 2000. Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, 3: 191–197.
- Gauthier, I. & M. J. Tarr 1997. Becoming a “Greeble” expert: exploring mechanisms for face recognition. *Vision Research*, 37: 1673–1682.
- Gauthier, I., P. Williams, M. J. Tarr & J. Tanaka 1998. Training “Greeble” experts: a framework for studying expert object recognition processes. *Vision Research*, 38: 2401–2428.
- Gazdar, G. & G. K. Pullum 1976. Truth-functional connectives in natural language. In S. S. Mufwene, C. A. Walker & S. B. Steever (eds) *Papers from the Twelfth Regional Meeting, Chicago Linguistic Society*, 220–234. Chicago: Chicago Linguistic Society.
- Gelman, R., F. Durgin & L. Kaufman 1995. Distinguishing between animates and inanimates: not by motion alone. In D. Sperber, D. Premack & A. J. Premack (eds) *Causal Cognition: A Multidisciplinary Debate*, 150–184. Oxford: Oxford University Press.
- Gentzen, G. 1935/1969. Investigations into logical deduction. In M. E. Szabo (ed.) *The Collected Papers of Gerhard Gentzen*, 68–131. Amsterdam: North Holland.
- Geurts, B. 1997. Good news about the description theory of names. *Journal of Semantics*, 14: 319–348.
- Gibson, E. J. 1969. *Principles of Perceptual Learning and Development*. New York: Appleton-Century-Crofts.

- Gibson, E. J., A. Pick, H. Osser & M. Hammond 1962. The role of grapheme-phoneme correspondence in the perception of words. *American Journal of Psychology*, 75: 554–570.
- Gibson, E. J. & R. D. Walk 1960. The “visual cliff”. *Scientific American*, 202.4 (April 1960): 64–71.
- Gibson, J. J. 1948. The Possibilities of Research on Training in Recognition and other Visual Performances. Report to Armed Forces by the NRC Vision Committee, University of Michigan.
- Gibson, J. J. 1966. *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin.
- Gibson, J. J. 1979. *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.
- Gibson, J. J. & R. M. Gagné 1947. Research on the recognition of aircraft. In J. J. Gibson (ed.) *Motion Picture Testing and Research*, 113–168. (Army Air Force Aviation Psychology Program Research Report No. 7.) Washington, DC: U.S. Government Printing Office.
- Gigerenzer, G., P. M. Todd & The ABC Research Group 1999. *Simple Heuristics That Make Us Smart*. Oxford: Oxford University Press.
- Gilbert, C. D. & T. N. Wiesel 1992. Receptive field dynamics in adult primary visual cortex. *Nature*, 356: 150–152.
- Goddard, C. & A. Wierzbicka (eds) 2002. *Meaning and Universal Grammar: Theory and Empirical Findings* (2 volumes). Amsterdam: John Benjamins.
- Goldstone, R. L. 1994. Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, 123: 178–200.
- Goldstone, R. L. 1998. Perceptual learning. *Annual Review of Psychology*, 49: 585–612.
- Goodman, N. 1976. *Languages of Art*, 2<sup>nd</sup> edition. Indianapolis: Hackett.

- Gould, J. L. 2002. Learning instincts. In H. Pashler & R. Gallistel (eds) *Stevens' Handbook of Experimental Psychology*, vol. 3, 3<sup>rd</sup> edition, 239–257. New York: Wiley.
- Gould, J. L. & P. Marler 1984. Ethology and the natural history of learning. In P. Marler & H. Terrace (eds) *The Biology of Learning*, 47–74. Berlin: Springer.
- Gould, J. L. & P. Marler 1987. Learning by instinct. *Scientific American*, 256.1 (January 1987): 74–85.
- Grice, H. P. 1957. Meaning. *Philosophical Review*, 66: 377–88.
- Grice, H. P. 1967. *Logic and Conversation*. William James Lectures, reprinted in Grice 1989, 1–143.
- Grice, H. P. 1989. *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.
- Gunther, Y. H. (ed.) 2003. *Essays on Nonconceptual Content*. Cambridge, MA: MIT Press.
- Hadjichristidis, C., R. J. Stevenson, D. E. Over, S. A. Sloman, J. St. B. T. Evans & A. Feeney 2001. On the evaluation of 'if p then q' conditionals. In J. D. Moore & K. Stenning (eds) *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society*, 409–414. London: Lawrence Erlbaum.
- Haig, N. D. 1984. The effect of feature displacement on face recognition. *Perception*, 13: 505–512.
- Hancock, P. J. B., V. Bruce & A. M. Burton 2000. Recognition of unfamiliar faces. *Trends in Cognitive Sciences*, 4: 330–337.
- Hanson, W. H. 1997. The concept of logical consequence. *The Philosophical Review*, 106: 365–409.
- Harman, G. H. 2002. Internal critique of logic and practical reasoning. In D. M. Gabbay & H. J. Ohlbach (eds) *Studies in Logic and Practical Reasoning*, vol. I, 171–186. Amsterdam: Elsevier.
- Harnad, S. 1996. Experimental analysis of naming behavior cannot explain naming capacity. *Journal of the Experimental Analysis of Behavior*, 65: 262–264.

- Heiling, A. M. & M. E. Herberstein 1999. The role of experience in web-building spiders (Araneidae). *Animal Cognition*, 2: 171–177.
- Hermer, L. & E. Spelke 1996. Modularity and development: the case of spatial reorientation. *Cognition*, 61: 195–232.
- Hillis, A. & A. Caramazza 1991. Category-specific naming and comprehension impairment: a double dissociation. *Brain*, 114: 2081–2094.
- Hirschfeld, L. 2001. On a folk theory of society: children, evolution, and mental representations of social groups. *Personality and Social Psychology Review*, 5: 107–117.
- Hochstein, S. & M. Ahissar 2002. View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron*, 36: 791–804.
- Hoffman, D. D. & M. Singh 1997. Saliency of visual parts. *Cognition*, 63: 29–78.
- Horton, M. S. & E. M. Markman 1980. Developmental differences in the acquisition of basic and superordinate categories. *Child Development*, 51: 708–719.
- Horwich, P. 1992. Chomsky versus Quine on the analytic–synthetic distinction. *Proceedings of the Aristotelian Society*, 92: 95–108.
- Horwich, P. 1998. *Meaning*. Oxford: Oxford University Press.
- Hubel, D. H. & T. N. Wiesel 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology (London)*, 160: 106–154.
- Hubel, D. H. & T. N. Wiesel 1968. Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology (London)*, 195: 215–243.
- Hugson, A. & R. Boakes 2001. Perceptual and cognitive elements of wine tasting expertise. *Australian Journal of Psychology*, 53: 103–108.
- Hurford, J. R. 1974. Exclusive or inclusive disjunction. *Foundations of Language*, 11: 409–411.
- Intraub, H. 1981. Rapid conceptual identification of sequentially presented pictures. *Journal of Experimental Psychology: Human Perception and Performance*, 7: 604–610.

- Jackendoff, R. 1992. *Languages of the Mind: Essays on Mental Representation*.  
Cambridge, MA: MIT Press.
- Jackson, F. 1998. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*.  
Oxford: Oxford University Press.
- Jaggi, S., W. C. Karl, S. G. Mallat & A. S. Willsky 1999. Silhouette recognition  
using high-resolution pursuit. *Pattern Recognition*, 32: 753–771.
- James, W. 1890. *The Principles of Psychology*, vol. 1. New York: Holt.
- Jennings, R. E. 1994. *The Genealogy of Disjunction*. Oxford: Oxford University  
Press.
- Jiang, Y. & M. M. Chun 2001. Selective attention modulates implicit learning.  
*Quarterly Journal of Experimental Psychology: Human Experimental  
Psychology*, 54A: 1105–1124.
- Johnson, K. E. 1992. *The Effect of Expertise on Hierarchical Systems of  
Categorization*. Ph.D. Dissertation, Emory University, Atlanta.
- Johnson, K. E. & C. B. Mervis 1997. Effects of varying levels of expertise on the  
basic level of categorization. *Journal of Experimental Psychology: General*,  
126: 248–277.
- Johnson-Laird, P. N. 1993. *Human and Machine Thinking*. Hillsdale, NJ: Erlbaum.
- Johnson-Laird, P. N. 2005. If bears eat in the woods...? Review of Evans & Over  
(2004). *Trends in Cognitive Sciences*, 9: 43–44.
- Johnson-Laird, P. N. & J. Tagart 1969. How implication is understood. *American  
Journal of Psychology*, 82: 367–373.
- Jones, S. 1970. Visual and verbal processes in problem-solving. *Cognitive  
Psychology*, 1: 201–214.
- Jones, S. S. & L. B. Smith 2002. How children know the relevant properties for  
generalizing object names. *Developmental Science*, 5: 219–232.
- Jones, S. S., L. B. Smith & B. Landau 1991. Object properties and knowledge in  
early lexical learning. *Child Development*, 62: 499–516.

- Jubien, M. 1989. On properties and property theory. In G. Chierchia, B. H. Partee & R. Turner (eds) *Properties, Types and Meaning*, vol. 1, 159–175. Dordrecht: Kluwer.
- Kalderon, M. E. 2001. Reasoning and representing. *Philosophical Studies*, 105: 129–160.
- Karni, A. & D. Sagi 1993. The time course of learning a visual skill. *Nature*, 365: 250–252.
- Katz, J. 1972. *Semantic Theory*. New York: Harper & Row.
- Keil, F. C. & R. A. Wilson 2000. The concept concept: the wayward path of cognitive science. *Mind and Language*, 15: 308–318.
- Klahr, D. 1973. A production system for counting, subitizing and adding. In W. G. Chase (ed.) *Visual Information Processing*, 527–546. New York: Academic Press.
- Kneale, W. 1962. Modality de dicto and de re. In E. Nagel, P. Suppes & A. Tarski (eds) *Logic, Methodology and Philosophy of Science. Proceedings of the 1960 International Congress*, 622–633. Stanford: Stanford University Press.
- Kobatake, E. & K. Tanaka 1994. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology*, 71: 856–867.
- Koslow, A. 1992. *A Structuralist Theory of Logic*. Cambridge: Cambridge University Press.
- Kreiman, G., C. Koch & I. Fried 2000. Category-specific visual responses of single neurons in the human medial temporal lobe. *Nature Neuroscience*, 3: 946–953.
- Kripke, S. 1972/1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- Lakoff, G. & J. R. Ross 1976. Is deep structure necessary? In J. D. McCawley (ed.) *Syntax and Semantics, vol. 7: Notes from the Linguistic Underground*, 159–164. New York: Academic Press.



- Lakoff, R. 1971. If's, and's and but's about conjunction. In C. J. Fillmore & D. T. Langendoen (eds) *Studies in Linguistic Semantics*, 114–149. New York: Holt, Rinehart & Winston.
- Landau, B. 1994. Object shape, object name, and object kind. *The Psychology of Learning and Motivation*, 31: 253–304.
- Landau, B. 2000. Concepts, the lexicon and acquisition: Fodor's new challenge. *Mind and Language*, 15: 319–326.
- Landau, B., L. Smith & S. Jones 1998a. Object perception and object naming in early development. *Trends in Cognitive Sciences*, 2: 19–24.
- Landau, B., L. Smith & S. Jones 1998b. Object shape, object function, and object naming. *Journal of Memory and Language*, 38: 1–27.
- Laurence, S. & E. Margolis 1999a. Concepts and cognitive science. In Margolis & Laurence (eds) *Concepts: Core Readings*, 3–81. Cambridge, MA: MIT Press.
- Laurence, S. & E. Margolis 1999b. Review of Fodor (1998a). *British Journal for the Philosophy of Science*, 50: 487–491.
- Laurence, S. & E. Margolis 2003a. Should we trust our intuitions? *Proceedings of the Aristotelian Society*, 103: 299–323.
- Laurence, S. & E. Margolis 2003b. Concepts. In S. P. Stich & T. Warfield (eds) *The Blackwell Guide to the Philosophy of Mind*, 190–213. Oxford, Blackwell.
- Lepore, E. 2000. *Meaning and Argument*. Oxford: Blackwell.
- Leslie, A. M. 1994. Pretending and believing: issues in the theory of ToMM. *Cognition*, 50: 211–238.
- Leslie, A. M., O. Friedman & T. P. German 2004. Core mechanisms in 'theory of mind'. *Trends in Cognitive Sciences*, 8: 528–533.
- Leslie, A. M. & S. Keeble 1987. Do 6-month old infants perceive causality? *Cognition*, 25: 265–288.
- Levine, A. & M. H. Bickhard 1999. Concepts: Where Fodor went wrong. *Philosophical Psychology*, 12: 5–23.

- Lewicki, P., T. Hill & M. Czyzewska 1992. Nonconscious acquisition of information. *American Psychologist*, 47: 796–801.
- Lieberman, M. D., R. Gaunt, D. T. Gilbert & Y. Trope 2002. Reflexion and reflection: a social cognitive neuroscience approach to attributional inference. *Advances in Experimental Social Psychology*, 34: 199–249.
- Loewer, B. & G. Rey (eds) 1991. *Meaning in Mind: Fodor and his Critics*. Oxford: Blackwell.
- Łukasiewicz, J. 1920/1970. On three-valued logic. In Łukasiewicz 1970, 87–88.
- Łukasiewicz, J. 1970. *Selected Works* (ed. L. Borkowski). Amsterdam: North Holland.
- Lukatela, G. & M. T. Turvey 1991. Phonological access of the lexicon: evidence from associative priming with pseudohomophones. *Journal of Experimental Psychology: Human Perception and Performance*, 17: 951–966.
- Lukatela, G. & M. T. Turvey 1993. Similar attentional, frequency, and associative effects for pseudohomophones and words. *Journal of Experimental Psychology: Human Perception and Performance*, 19: 166–178.
- Lukatela, G. & M. T. Turvey 1994a. Visual lexical access is initially phonological: 1. Evidence from associative priming by words, homophones, and pseudohomophones. *Journal of Experimental Psychology: General*, 123: 107–128.
- Lukatela, G. & M. T. Turvey 1994b. Visual lexical access is initially phonological: 2. Evidence from phonological priming by homophones and pseudohomophones. *Journal of Experimental Psychology: General*, 123: 331–353.
- Lunn, J. H. 1948. Chick sexing. *American Scientist*, 36: 280–287.
- Lycan, W. 1984. *Logical Form in Natural Language*. Cambridge, MA: MIT Press.
- Lyons, J. 1977. *Semantics* (2 volumes). Cambridge: Cambridge University Press.

- McCandliss, B. D., L. Cohen & S. Dehaene 2003. The visual word form area: expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences*, 7: 293–299.
- McCawley, J. D. 1968. Lexical insertion in a transformational grammar without deep structure. *Papers from the Fourth Regional Meeting of the Chicago Linguistic Society*, 71–80. Reprinted in J. D. McCawley 1976. *Grammar and Meaning*, 155–166. New York: Academic Press.
- McGinn, C. 1989. *Mental Content*. Oxford: Blackwell.
- Mack, A. 2003. Inattention blindness: looking without seeing. *Current Directions in Psychological Science*, 12: 180–184.
- Mack, A. & I. Rock 1998. *Inattention Blindness*. Cambridge, MA: MIT Press.
- Madole, K. L. & L. B. Cohen 1995. The role of object parts in infants' attention to form–function correlations. *Developmental Psychology*, 31: 637–648.
- Malt, B. C. & E. C. Johnson 1998. Artifact category membership and the intentional-historical theory. *Cognition*, 66: 79–85.
- Margolis, E. 1998. How to acquire a concept. *Mind and Language*, 13: 347–369.
- Marler, P. 1991. The instinct to learn. In S. Carey & R. Gelman (eds) *The Epigenesis of Mind: Essays on Biology and Cognition*, 37–66. London: Lawrence Erlbaum.
- Marler, P. 2004. Innateness and the instinct to learn. *Anais da Academia Brasileira de Ciências*, 76: 189–200.
- Marr, D. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: Freeman.
- Martin, J. N. 1987. *Elements of Formal Semantics*. London: Academic Press.
- Martin, R. D. 1994. *The Specialist Chick Sexer: A History, A World View, Future Prospects*. Melbourne: Bernal Publishing.
- Masui, K. & J. Hashimoto 1933. *Sexing Baby Chicks*. Vancouver: Journal Printing Company.
- Matan, A. & S. Carey 2001. Developmental changes within the core of artifact concepts. *Cognition*, 78: 1–26.

- Maye, J., J. F. Werker & L. Gerken 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82: B101–B111.
- Miller, G. A. 1978. Semantic relations among words. In M. Halle, J. Bresnan & G. Miller (eds) *Linguistic Theory and Psychological Reality*, 60–118. Cambridge, MA: MIT Press.
- Millikan, R. G. 1984. *Language, Thought, and Other Biological Categories: New Foundations for Realism*. Cambridge, MA: MIT Press.
- Millikan, R. G. 1993. On Mentalese orthography. In B. Dahlbom (ed.) *Dennett and his Critics: Demystifying Mind*, 97–123. Oxford: Blackwell.
- Milne, P. 1994. Classical harmony: rules of inference and the meaning of the logical constants. *Synthese*, 100: 49–94.
- Morton, J. & M. H. Johnson 1991. CONSPEC and CONLERN: a two-process theory of infant face recognition. *Psychological Review*, 98: 164–181.
- Murphy, G. L. 2002. *The Big Book of Concepts*. Cambridge, MA: MIT Press.
- Musgrave, A. 1993. *Common Sense, Science and Scepticism: A Historical Introduction to the Theory of Knowledge*. Cambridge: Cambridge University Press.
- Myers, D. G. 2002. *Intuition: Its Powers and Perils*. New Haven, CT: Yale University Press.
- Nelson, C. A. 2001. The development and neural bases of face recognition. *Infant and Child Development*, 10: 3–18.
- Nelson, M. 2002. Descriptivism defended. *Noûs*, 36: 408–436.
- Neuner, F. & S. R. Schweinberger 2000. Neuropsychological impairments in the recognition of faces, voices, and personal names. *Brain and Cognition*, 44: 342–366.
- Nichols, S. & S. Stich 2000. A cognitive theory of pretense. *Cognition*, 74: 115–147.
- Noveck, I. A. 2004. Pragmatic inferences linked to logical terms. In I. A. Noveck & D. Sperber (eds) *Experimental Pragmatics*, 301–321. Basingstoke, UK: Palgrave Macmillan.

- Opfer, J. E. & R. S. Siegler 2004. Revisiting preschoolers' *living things* concept: a microgenetic analysis of conceptual change in basic biology. *Cognitive Psychology*, 49: 301–332.
- Over, D. E. & J. St. B. T. Evans 2003. The probability of conditionals: The psychological evidence. *Mind & Language*, 18: 340–358.
- Palmeri, T. J. 1997. Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23: 324–354.
- Partee, B. H., A. ter Meulen & R. E. Wall 1993. *Mathematical Methods in Linguistics*. Dordrecht: Kluwer.
- Pascalis, O., E. Demont, M. de Haan & R. Campbell 2001. Recognition of faces of different species: a developmental study between 5 and 8 years of age. *Infant and Child Development*, 10: 39–45.
- Pascalis, O., S. de Schonen, J. Morton, C. Deruelle & M. Fabre-Grenet 1995. Mother's face recognition by neonates: a replication and an extension. *Infant Behavior and Development*, 18: 79–85.
- Patel, V. L., J. F. Arocha & D. R. Kaufmann 1994. Diagnostic reasoning and medical expertise. In D. Medin (ed.) *The Psychology of Learning and Motivation*, vol. 31, 187–251. New York: Academic Press.
- Pavlov, I. 1927. *Conditioned Reflexes*. Oxford: Oxford University Press.
- Peacocke, C. 1983. *Sense and Content*. Oxford: Oxford University Press.
- Peacocke, C. 1992. *A Study of Concepts*. Cambridge, MA: MIT Press.
- Peacocke, C. 2004. Interrelations: Concepts, knowledge, reference and structure. *Mind and Language*, 19: 85–98.
- Peirce, C. S. 1931–1935. *Collected Papers*, vol. II. Cambridge, MA: Harvard University Press.
- Pelletier, F. J. 1977. Or. *Theoretical Linguistics*, 4: 61–74.
- Perner, J. & B. Lang 2000. Theory of mind and executive function: is there a developmental relationship? In S. Baron-Cohen, H. Tager-Flusberg & D. J.

- Cohen (eds) *Understanding Other Minds*, 2<sup>nd</sup> edition (Perspectives from Developmental Cognitive Neuroscience), 150–181. Oxford: Oxford University Press.
- Piattelli-Palmarini, M. 1989. Evolution, selection and cognition: from “learning” to parameter setting in biology and in the study of language. *Cognition*, 31: 1–44.
- Premack, D. 1990. The infant’s theory of self-propelled objects. *Cognition*, 36: 1–16.
- Premack, D. & A. J. Premack 1995. Intention as psychological cause. In D. Sperber, D. Premack & A. J. Premack (eds) *Causal Cognition: A Multidisciplinary Debate*, 185–199. Oxford: Oxford University Press.
- Prinz, J. & A. Clark 2004. Putting concepts to work: some thoughts for the twenty-first century. *Mind and Language*, 19: 57–69.
- Prior, A. N. 1960. The runabout inference-ticket. *Analysis*, 21: 38–39.
- Pullum, G. K. 1991. The great Eskimo vocabulary hoax. In *The Great Eskimo Vocabulary Hoax and Other Irreverent Essays on the Study of Language*, 159–171. Chicago: University of Chicago Press.
- Putnam, H. 1975. The meaning of “meaning”. In K. Gunderson (ed.) *Minnesota Studies in the Philosophy of Science, Volume VII: Language, Mind, and Knowledge*, 131–193. Minneapolis: University of Minnesota Press.
- Putnam, H. 1983. ‘Two dogmas’ revisited. In *Realism and Reason (Philosophical Papers, vol. 3)*, 87–97. Cambridge: Cambridge University Press.
- Pylyshyn, Z. W. 2000. Situating vision in the world. *Trends in Cognitive Sciences*, 4: 197–207.
- Pyysiäinen, I. 2003. Dual-process theories and hybrid systems. *Behavioral and Brain Sciences*, 26: 617–618. (Commentary on Anderson & Lebiere 2003.)
- Quine, W. V. O. 1935/1976. Truth by convention. In Quine 1976, 77–106.
- Quine, W. V. O. 1948/1953. On what there is. In Quine 1953b, 1–19.
- Quine, W. V. O. 1953a. Two dogmas of empiricism. In Quine 1953b, 20–46
- Quine, W. V. O. 1953b. *From a Logical Point of View*. Cambridge, MA: Harvard University Press.

- Quine, W. V. O. 1954/1976. Carnap and logical truth. In Quine 1976, 107–132.
- Quine, W. V. O. 1960. *Word and Object*. Cambridge, MA: MIT Press.
- Quine, W. V. O. 1973/1976. Vagaries of definition. In Quine 1976, 50–55.
- Quine, W. V. O. 1976. *The Ways of Paradox and Other Essays*, revised and enlarged edition. Cambridge, MA: Harvard University Press.
- Quine, W. V. O. 1977. Natural kinds. In S. P. Schwartz (ed.) *Naming, Necessity, and Natural Kinds*, 155–175. Ithaca: Cornell University Press.
- Rakison, D. H. & G. E. Butterworth 1998. Infants' use of object parts in early categorization. *Developmental Psychology*, 34: 49–62.
- Ray, G. 1997. Fodor and the inscrutability problem. *Mind and Language*, 12: 475–489.
- Rayner, K., B. R. Foorman, C. A. Perfetti, D. Pesetsky & M. S. Seidenberg 2002. How should reading be taught? *Scientific American*, 286.3 (March 2002): 70–77.
- Reed, C. F., P. N. Witt, M. B. Scarborough & D. B. Peakall 1970. Experience and the orb web. *Developmental Psychobiology*, 3: 251–265.
- Reichenbach, H. 1944. *Philosophical Foundations of Quantum Mechanics*. Berkeley: University of California Press.
- Rhodes, G. 1988. Looking at faces: first-order and second-order features as determinants of facial appearance. *Perception*, 17: 43–63.
- Ridley, M. 2003. *Nature Via Nurture: Genes, Experience, and What Makes Us Human*. London: Fourth Estate.
- Rips, L. 1983. Cognitive processes in propositional reasoning. *Psychological Review*, 90: 38–71.
- Robbins, R. & E. McKone 2003. Can holistic processing be learned for inverted faces? *Cognition*, 88: 79–107.
- Rosch, E., C. B. Mervis, W. Gray, D. Johnson & P. Boyes-Braem 1976. Basic objects in natural categories. *Cognitive Psychology*, 8: 382–439.

- Rossion, B., L. Dricot, A. Devolder, J.-M. Bodart, M. Crommelinck, B. De Gelder & R. Zoontjes 2000. Hemispheric asymmetries for whole-based and part-based face processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience*, 12: 793–802.
- Sacchett, C. & G. W. Humphreys 1992. Calling a squirrel a squirrel but a canoe a wigwam: a category-specific deficit for artefactual objects and body parts. *Cognitive Neuropsychology*, 9: 73–86.
- Samuels, R. 2002. Nativism in cognitive science. *Mind and Language*, 17: 233–265.
- Samuels, R. 2004. Innateness in cognitive science. *Trends in Cognitive Sciences*, 8: 136–141.
- Samuels, R. 2005. The complexity of cognition: tractability arguments for massive modularity. In Carruthers et al. (eds) 2005, 107–121.
- Samuels, R. & S. Stich 2004. Rationality and psychology. In A. R. Mele & P. Rawling (eds) *The Oxford Handbook of Rationality*, 279–300. Oxford: Oxford University Press.
- Sapolsky, R. 2003. Bugs in the brain. *Scientific American*, 288.3: 94–97 (March 2003).
- Schiffer, S. 1981. Truth and the theory of content. In H. Parret & J. Bouvaresse (eds) *Meaning and Understanding*, 204–222. Berlin: Walter de Gruyter.
- Schneider, W. & J. M. Chein 2003. Controlled & automatic processing: behavior, theory, and biological mechanisms. *Cognitive Science*, 27: 525–559.
- Schyns, P. G., R. L. Goldstone & J.-P. Thibaut 1998. The development of features in object concepts. *Behavioral and Brain Sciences*, 21: 1–54.
- Schyns, P. G. & G. L. Murphy 1994. The ontogeny of part representation in object concepts. In D. Medin (ed.) *The Psychology of Learning and Motivation*, vol. 31, 305–354. New York: Academic Press.
- Schyns, P. G. & L. Rodet 1997. Categorization creates functional features. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23: 681–696.



- Scott, S. 2002. The psychological implausibility of naturalized content. In W. D. Gray & C. D. Schunn (eds) *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society*, 822–827. Mahwah, NJ: Lawrence Erlbaum.
- Scott, S. 2003. *Non-Referring Concepts*. Ph.D. Dissertation, Carleton University, Ottawa. [Carleton University Cognitive Science Technical Report 2003-03. URL = <<http://www.carleton.ca/iis/TechReports>>]
- Searcy, J. H. & J. C. Bartlett 1996. Inversion and processing of component and spatial-relational information in faces. *Journal of Experimental Psychology: Human Perception and Performance*, 22: 904–915.
- Segal, G. M. A. 2000. *A Slim Book about Narrow Content*. Cambridge, MA: MIT Press.
- Sergent, J. 1984. An investigation into component and configural processes underlying face perception. *British Journal of Psychology*, 75: 221–242.
- Sewards, T. V. & M. A. Sewards 2002. Innate visual object recognition in vertebrates: some proposed pathways and mechanisms. *Comparative Biochemistry and Physiology Part A*, 132: 861–891.
- Shastri, L. & V. Ajjanagadde 1993. From simple associations to systematic reasoning: a connectionist representation of rules, variables and dynamic bindings using temporal synchrony. *Behavioral and Brain Sciences*, 16: 417–494.
- Sher, G. 1991. *The Bounds of Logic: A Generalized Viewpoint*. Cambridge, MA: MIT Press.
- Sher, G. 1996. Semantics and Logic. In S. Lappin (ed.) *The Handbook of Contemporary Semantic Theory*, 511–537. Oxford: Blackwell.
- Simon, H. A. 1992. What is an “explanation” of behavior? *Psychological Science*, 3: 150–161.
- Simon, H. A. & W. G. Chase 1973. Skill in chess. *American Scientist*, 61: 394–403.

- Simons, D. J. & C. F. Chabris 1999. Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception*, 28: 1059–1074.
- Sloman, S. A. 1996. The empirical case for two systems of reasoning. *Psychological Bulletin*, 119: 3–22.
- Smith, E. R. & J. DeCoster 2000. Dual-process models in social and cognitive psychology: conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, 4: 108–131.
- Snowling, M. J. & C. Hulme (eds) 2005. *The Science of Reading: A Handbook*. Oxford: Blackwell.
- Sober, E. 1982. Why logically equivalent predicates may pick out different properties. *American Philosophical Quarterly*, 19: 183–189.
- Solomon, G. E. A. 1990. Psychology of novice and expert wine talk. *American Journal of Psychology*, 103: 495–517.
- Sosa, D. 2001. Rigidity in the scope of Russell's theory. *Noûs*, 35: 1–38.
- Sperber, D. 1975. *Rethinking Symbolism*. Cambridge: Cambridge University Press.
- Sperber, D. 1994. The modularity of thought and the epidemiology of representations. In L. A. Hirschfeld & S. A. Gelman (eds) *Mapping the Mind: Domain Specificity in Cognition and Culture*, 39–67. Cambridge: Cambridge University Press.
- Sperber, D. 1996. *Explaining Culture: A Naturalistic Approach*. Oxford: Blackwell.
- Sperber, D. 1997. Intuitive and reflective beliefs. *Mind and Language*, 12: 67–83.
- Sperber, D. (ed.) 2000. *Metarepresentations: A Multidisciplinary Perspective*. Oxford: Oxford University Press.
- Sperber, D. 2005. Modularity and relevance: How can a massively modular mind be flexible and context-sensitive? In Carruthers et al. (eds) 2005, 53–68.
- Sperber, D., F. Cara & V. Girotto 1995. Relevance theory explains the selection task. *Cognition*, 52: 3–22.
- Sperber, D. & D. Wilson 1983. Draft of *Relevance*. Ms., UCL.
- Sperber, D. & D. Wilson 1995. *Relevance*, 2<sup>nd</sup> edition. Oxford: Blackwell.

- Sperber, D. & D. Wilson 1998. The mapping between the mental and the public lexicon. In P. Carruthers & J. Boucher (eds) *Language and Thought*, 184–200. Cambridge: Cambridge University Press.
- Sperber, D. & D. Wilson 2002. Pragmatics, modularity and mind-reading. *Mind and Language*, 17: 3–23.
- Stainton, R. J. & C. Viger 2000. Review of Fodor (1998a). *Synthese*, 123: 131–151.
- Stalnaker, R. 1998. What might nonconceptual content be? *Philosophical Issues*, 9: 339–352. Reprinted in Gunther 2003, 95–106.
- Stanley, J. 1997. Names and rigid designation. In B. Hale & C. Wright (eds) *A Companion to the Philosophy of Language*, 555–585. Oxford: Blackwell.
- Stich, S. P. 1978. Beliefs and subdoxastic states. *Philosophy of Science*, 45: 499–518.
- Stroop, J. R. 1935. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 12: 242–248.
- Swoyer, C. 1996. Theories of properties: from plenitude to paucity. *Philosophical Perspectives*, 10: 243–264 (Supplement to *Noûs*, volume 30).
- Swoyer, C. 2000. Properties. In E. N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy (Winter 2000 Edition)*, URL = <<http://plato.stanford.edu/archives/win2000/entries/properties/>>.
- Tanaka, J. W. & T. Curran 2001. A neural basis for expert object recognition. *Psychological Science*, 12: 43–47.
- Tanaka, J. W., T. Curran & D. L. Sheinberg 2005. The training and transfer of real-world perceptual expertise. *Psychological Science*, 16: 145–151.
- Tanaka, J. W. & M. J. Farah 1993. Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 46A: 225–245.
- Tanaka, J. W., M. Kiefer & C. Bukach 2004. A holistic account of the own-race effect in face recognition: evidence from a cross-cultural study. *Cognition*, 93: B1–B9.

- Tanaka, J. W. & J. A. Sengco 1997. Features and their configuration in face recognition. *Memory & Cognition*, 25: 583–592.
- Tanaka, J. W. & M. Taylor 1991. Object categories and expertise: is the basic level in the eye of the beholder? *Cognitive Psychology*, 23: 457–82.
- Tarski, A. 1936/1956. On the concept of logical consequence. In Tarski 1956, 409–420.
- Tarski, A. 1956. *Logic, Semantics, Metamathematics* (translated by J. H. Woodger). Oxford: Oxford University Press.
- Tarski, A. 1966/1986. What are logical notions? *History and Philosophy of Logic*, 7: 143–154. (Text of a 1966 lecture, edited by J. Corcoran.)
- Tennant, N. 1987. *Anti-realism and Logic*. Oxford: Oxford University Press.
- Thorpe, S., D. Fize & C. Marlot 1996. Speed of processing in the human visual system. *Nature*, 381: 520–522.
- Tinbergen, N. 1951. *The Study of Instinct*. Oxford: Oxford University Press.
- Tooby, J., L. Cosmides & C. Barrett 2005. Resolving the debate on innate ideas: Learnability constraints and the evolved interpenetration of motivational and conceptual functions. In Carruthers et al. (eds) 2005, 305–337.
- Treisman, A. & G. Gelade 1980. A feature-integration theory of attention. *Cognitive Psychology*, 12: 97–136.
- Trick, L. M. & Z. W. Pylyshyn 1993. What enumeration studies can show us about spatial attention: evidence for limited capacity preattentive processing. *Journal of Experimental Psychology: Human Perception and Performance*, 19: 331–351.
- Tulving, E. 1972. Episodic and semantic memory. In E. Tulving & W. Donaldson (eds) *Organization of Memory*, 381–403. New York: Academic Press.
- Tversky, A. 1972. Elimination by aspects: a theory of choice. *Psychological Review*, 79: 281–299.
- United States Army. 1996. *Visual Aircraft Recognition*. Field Manual No. 44-80 (Rev. 30 September 1996). Washington, DC: Department of the Army.

- Vallentyne, P. 1998. The nomic role account of carving reality at the joints. *Synthese*, 115: 171–198.
- Van Dijk, T. A. 1977. *Text and Context: Explorations in the Semantics and Pragmatics of Discourse*. London: Longman.
- Van Orden, G. C. 1987. A ROWS is a ROSE: spelling, sound, and reading. *Memory & Cognition*, 15: 181–198.
- Van Orden, G. C. 1991. Phonologic mediation is fundamental to reading. In Besner & Humphreys (eds) 1991, 77–103.
- Van Orden, G. C., J. C. Johnston & B. L. Hale 1988. Word identification in reading proceeds from spelling to sound to meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14: 371–386.
- Van Orden, G. C. & H. Kloos 2005. The question of phonology and reading. In Snowling & Hulme (eds) 2005, 61–78.
- Varzi, A. C. 2002. On logical relativity. *Philosophical Issues*, 10: 197–219.
- Viger, C. D. 2001. Locking on to the language of thought. *Philosophical Psychology*, 14: 203–215.
- von Kriegstein, K., A. Kleinschmidt, P. Sterzer & A.-L. Giraud 2005. Interaction of face and voice areas during speaker recognition. *Journal of Cognitive Neuroscience*, 17: 367–376.
- Wachowitz, S. & J.-P. Ewert 1996. A key by which the toad's visual system gets access to the domain of prey. *Physiology & Behavior*, 60: 877–887.
- Wakefield, J. C. 2003. Fodor on inscrutability. *Mind and Language*, 18: 524–537.
- Warrington, E. K. & T. Shallice 1984. Category specific semantic impairments. *Brain*, 107: 829–854.
- Wason, P. C. 1966. Reasoning. In B. M. Foss (ed.) *New Horizons in Psychology*, vol. 1, 135–151. Harmondsworth: Penguin.
- Watanabe, T., J. E. Náñez & Y. Sasaki 2001. Perceptual learning without perception. *Nature*, 413: 844–848.
- Watson, J. B. 1925. *Behaviorism*. New York: People's Institute.

- Wellman, H. M. & S. A. Gelman 1998. Knowledge acquisition in foundational domains. In D. Kuhn & R. S. Siegler (eds) *Handbook of Child Psychology*, vol. 2, 5<sup>th</sup> edition, 523–573. New York: Wiley.
- Werker, J. F. & R. C. Tees 1984. Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7: 49–63.
- Whiten, A. 2000. Primate culture and social learning. *Cognitive Science*, 24: 477–508.
- Whiten, A., J. Goodall, W. C. McGrew, T. Nishida, V. Reynolds, Y. Sugiyama, C. E. G. Tutin, R. W. Wrangham & C. Boesch 1999. Cultures in chimpanzees. *Nature*, 399: 682–685.
- Wierzbicka, A. 1972. *Semantic Primitives* (translated by A. Wierzbicka & J. Besemeres). Frankfurt: Athenäum.
- Williams, P., I. Gauthier & M. J. Tarr 1998. Feature learning during the acquisition of perceptual expertise. (Commentary on Schyns et al. 1998.) *Behavioral and Brain Sciences*, 21: 40–41.
- Wilson, D. 2000. Metarepresentation in linguistic communication. In Sperber (ed.) 2000, 411–448.
- Wolfe, J. M. 2003. Moving towards solutions to some enduring controversies in visual search. *Trends in Cognitive Sciences*, 7: 70–76.
- Woodfield, A. 1986. Two categories of content. *Mind and Language*, 1: 319–354.
- Young, A. W., D. Hallowell & D. C. Hay 1987. Configurational information in face perception. *Perception*, 16: 747–759.

*While I, no doubt, have learnt something extra about chick sexing,  
I was also reminded that man's abilities are limited only by his  
thinking.*

—R. D. Martin in his epilogue to *The Specialist Chick Sexer*