Intuition as Evidence in Philosophical Analysis:
Taking Connectionism Seriously
Ph.D. Degree, 2008, by Tom Rand, Graduate Department of Philosophy, University of Toronto

ABSTRACT

1. Intuitions are often treated in philosophy as a basic evidential source to confirm/discredit a proposed definition or theory; e.g. intuitions about Gettier cases are taken to deny a justified-true-belief analysis of ‘knowledge’. Recently, Weinberg, Nichols & Stitch (WN&S) provided evidence that epistemic intuitions vary across persons and cultures. In-so-far as philosophy of this type (Standard Philosophical Methodology – SPM) is committed to provide conceptual analyses, the use of intuition is suspect – it does not exhibit the requisite normativity. I provide an analysis of intuition, with an emphasis on its neural – or connectionist – cognitive backbone; the analysis provides insight into its epistemic status and proper role within SPM. Intuition is initially characterized as the recognition of a pattern.

2. The metaphysics of ‘pattern’ is analyzed for the purpose of denying that traditional symbolic computation is capable of differentiating the patterns of interest.

3. The epistemology of ‘recognition’ is analyzed, again, to deny that traditional computation is capable of capturing human acts of recognition.

4. Fodor’s informational semantics, his Language of Thought and his Representational Theory of Mind are analyzed and his arguments denied. Again, the purpose is to deny traditional computational theories of mind.

5. Both intuition and a theory of concepts – pragmatic conceptualism - are developed within the connectionist computational paradigm. Intuition is a particular sort of occurrent signal, and a concept is a counterfactually defined set of signals. Standard connectionist theory is significantly extended to develop my position, and consciousness plays a key functional role. This extension – taking connectionism seriously – is argued to be justified on the basis of the failure of the traditional computing paradigm to account for human cognition.

6. Repercussions for the use of intuition in SPM are developed. Variance in intuition is characterized – and expected - as a kind of bias in the network, either inherent or externally-provoked. The WN&S data is explained in the context of this bias. If SPM remains committed to the use of intuition, then intuition must be taken as a part of a larger body of evidence, and it is from experts – not the folk – that intuitions should be solicited.
What do we know about intuition? What idea have we of it? It’s presumably supposed to be a sort of seeing, recognition at a single glance; I wouldn’t know what more to say.¹

- Wittgenstein

Good Quinean that I am, I think that it is always up for grabs what an intuition is an intuition of.²

- Fodor

¹ Wittgenstein, Cause and Effect: Intuitive Awareness, 18:10
² Fodor, Concepts, pg., 87
TABLE OF CONTENTS

ABSTRACT .................................................................................................................. II

OVERVIEW .................................................................................................................. 1

ARGUMENT FORM .................................................................................................... 6

1. INTUITION ............................................................................................................. 7
   1.1 INTRODUCTION .................................................................................................. 7
   1.2 INTUITIONS ABOUT KNOWLEDGE ASCRIPTIONS: WEINBERG, NICHOLS AND STICH’S NORMATIVITY PROBLEM .......................................................... 11
      1.2.1 The Normative and Descriptive Projects ..................................................... 11
      1.2.2 Epistemic Romanticism & Intuition Driven Romanticism ......................... 12
      1.2.3 The Normativity Problem ........................................................................ 13
   1.3 FOLK INTUITIONS; EVIDENCE OF WHAT? HYPOTHESES VS. EVIDENCE ................................................................................................................. 16
   1.4 TWO CONTEMPORARY ANALYSES OF INTUITION; GOLDMAN & PUST, PUST. ........................................................................................................... 21
      1.4.1 Goldman & Pust; The ‘Reliabilist’ View .................................................... 21
      1.4.2 Pust ............................................................................................................... 25
         1.4.2.1 Particularist versus Generalist Intuitions .............................................. 25
         1.4.2.2 Narrow and Wide Reflective Equilibriums, General and Global Intuitionism ........................................................................................................ 26
         1.4.2.3 The Nature of Intuition as “Intellectual Seeming” ............................... 30
         1.4.2.4 Intuition and Belief .............................................................................. 31
         1.4.2.5 Intuition and Necessity ...................................................................... 33
   1.5 INTUITION, A DEFINITION ............................................................................. 35
      1.5.1 H-Intuition; Reliable Hunches .................................................................. 36
      1.5.2 O-Intuition; Intuiting the Obvious ............................................................. 40
      1.5.3 Introspective Support and Epistemic Necessity ....................................... 41
      1.5.4 Intuition with Large: H- and O-Intuitions as Pattern Recognition .......... 46

2. PATTERN METAPHYSICS .................................................................................. 48
   2.1 METAPHYSICS OF PATTERNS – THE LANDSCAPE .................................. 49
      2.1.1 A Quick Note on Physicalism .................................................................... 50
   2.2 DENNETT AND REAL PATTERNS .................................................................. 52
      2.2.1 Barcodes and Noise .................................................................................. 54
         2.2.1.1 The Indeterminacy of Signal Choice; Expediency, Interests & Abilities ........................................................................................................... 56
         2.2.1.2 The Indeterminacy of Underlying Explanatory Processes ................ 59
      2.2.2 Life, Chess and a Multiplicity of Perspectives ......................................... 60
      2.2.3 Lessons: The Indeterminacy of Real Patterns .......................................... 62
      2.2.4 Dennett; Cautions & Conclusions ......................................................... 64
         2.2.4.1 Support for Ontological and Epistemic Continuity ......................... 66
   2.3 HAUGELAND; CONTEXT-DEPENDENCE AND NORMATIVITY ............. 68
      2.3.1 Strong Context-Dependence .................................................................... 69
      2.3.2 Normativity, Objecthood and Discernability ......................................... 72
      2.3.3 Concluding Remarks: Are Constitutive Standards Recognition-Independent? ............................................................................................. 75
   2.4 DREYFUS; THE KINDS OF PATTERNS WE RECOGNIZE ......................... 78
      2.4.1 Three Kinds of Patterns .......................................................................... 78
      2.4.2 Facts, Patterns and Language: The Ontological Assumption ............... 82
   2.5 PATTERNS; A HIERARCHY OF COMPLEXITY ........................................ 85
      2.5.1 Amenability to Algorithmic Description ................................................. 85
      2.5.2 Mind-Dependency ..................................................................................... 89
3. PATTERN EPISTEMOLOGY OR ‘RECOGNITION’ .................................................. 98
  3.1 THE PATTERN-RECOGNITION STANCE AND A NEW COMPUTING PRIMITIVE .......................................................... 100
  3.2 (HUMAN) PATTERN-RECOGNITION TRAITS – OR – WHY WE ARE NOT COMPUTING .................................................. 102
    3.2.1 Fringe Consciousness ................................................................................................................................. 104
      3.2.1.1 Accumulated Past and Present Actions as Context .................................................................................. 106
      3.2.1.2 Heuristics Top to Bottom .......................................................................................................................... 108
      3.2.1.3 Ambiguity of sub-Pattern Identity .............................................................................................................. 109
    3.2.2 Ambiguity Tolerance ................................................................................................................................. 110
      3.2.2.1 An Infinity of Facts vs. Recognizing a Situation .......................................................................................... 111
      3.2.2.2 Ambiguity and Particular Situational Recognition .................................................................................. 116
    3.2.3 Essential/Inessential Discrimination ............................................................................................................ 119
    3.2.4 Perspicuous Grouping and the Pattern-Rec Stance .......................................................................................... 121

4. CONCEPTUAL ATOMISM – FODOR ON CONCEPTS ............................................................................... 124
  4.1 MOTIVATION AND THE PSYCHOLOGICAL EXPLANANDA ................................................................................................. 126
    4.1.1 Fodor as Coherentist ......................................................................................................................................... 126
    4.1.2 The Desiderata; Causality & Publicity .................................................................................................................. 127
    4.1.2.a Interlude: Publicity, Similarity and Identity .................................................................................................. 131
    4.1.3 The Desiderata: Compositionality, Systematicity & Productivity ............................................................................ 133
  4.2 REPRESENTATIONAL THEORY OF MIND (RTM) ................................................................................................. 137
    4.2.1 Nomic Explanation ........................................................................................................................................... 137
    4.2.2 Mental Representations Bear Original Intentional Content ................................................................................. 138
    4.2.3 Thinking Equals Computation .......................................................................................................................... 139
    4.2.4 Informational Semantics ...................................................................................................................................... 141
      4.2.4.1 Causal-cum-Nomological Relations and Semantic Access .............................................................................. 141
      4.2.4.2 Inductive Learning, Stereotypes and Innate Mechanisms .............................................................................. 143
      4.2.4.3 Differentiating Coextensive Concepts; It’s in the Head ................................................................................. 147
      4.2.4.4 What’s so Bad About Inferential Role Semantics ? ...................................................................................... 150
    4.2.5 RTM: Summary .................................................................................................................................................. 152
  4.3 CONCEPTUAL ATOMISM ........................................................................................................................................... 153
    4.3.1 MOPs Slice as Words .......................................................................................................................................... 154
    4.3.2 Intuitions about Conceptual Connectedness and Analytic Truths ........................................................................... 155
  4.4 ANALYSIS .................................................................................................................................................................. 157
    4.4.1 Denying the Compositionality, Systematicity & Productivity Desiderata; Trees and Minds ...................................... 158
    4.4.2 Refuting (the need for) Nomic Intentional Explanations ......................................................................................... 165
    4.4.3 Refuting (Mentally-Tokened) Publicity ................................................................................................................ 169
    4.4.3.a Denying the ‘Digital’ Publicity Requirement ..................................................................................................... 170
  4.5 A SYMPATHETIC READING OF ATOMISM AND INFORMATIONAL SEMANTICS ............................................. 173
    4.5.1 The Constitutive Nature of Experience ................................................................................................................... 173
    4.5.2 Mind-Dependency ................................................................................................................................................ 174
    4.5.3 Ontology ............................................................................................................................................................... 175
    4.5.4 Stereotypes ........................................................................................................................................................... 176
    4.5.5 Semantic Access and Innateness .......................................................................................................................... 176

5. THE CONNECTIONIST STORY; IMPLEMENTATION MATTERS ..................................................................... 179
  5.1 OVERVIEW; C-NETS, MINDS, CONCEPTS AND INTUITION ...................................................................................... 179
  5.2 THE C-NETS OF INTEREST ........................................................................................................................................... 184
    5.2.1 The Role of Theory or, Not Just Perception ........................................................................................................... 185
    5.2.2 The Importance of Nonlinearity ........................................................................................................................... 186
5.2.3 Supervised vs. Unsupervised Learning ................................................................. 187
5.3 WHY C-NETS ARE NOT TURING MACHINES .......................................................... 190
  5.3.1 Superpositional Storage and the Semantic Metric ................................................. 190
  5.3.2 Intrinsic (or Rampant) Context Sensitivity and Dynamic Structures .................. 193
  5.3.3 Strong Representational Change; Code and Process .......................................... 198
  5.3.4 Concluding Remarks .......................................................................................... 200
5.4 THAT C-NETS ARE NOT TURING MACHINE – WHAT Follows? ............................... 201
  5.4.1 Explanatory Inversion and Post-Hoc Analysis ....................................................... 201
  5.4.2 Prototypes & Statistics, Theory & Knowledge, Intuition .................................... 204
    5.4.2.1 Typicality Judgments; Prototype Theory & The Statistical Central Tendency .... 205
    5.4.2.2 The Role of Theory & Knowledge ................................................................. 211
    5.4.2.3 Explicit Knowledge Structures ...................................................................... 213
  5.4.3 Pragmatic Conceptualism: A Theory of Concepts ................................................. 215
    5.4.3.1 Pragmatic Conceptualism, or, Never the Same Concept Twice ...................... 216
    5.4.3.2 Unsatisfied/Further Explanatory Obligations ................................................. 225
  5.4.4 Occurrent Signals as Intuitive Judgments .............................................................. 230

6. INTUITION AS EVIDENCE REVISITED ..................................................................... 236
  6.1 ON THE PROPER ROLE OF INTUITION ................................................................. 238
    6.1.1 The Intuition Skeptics ...................................................................................... 238
    6.1.2 Intuition; Ubiquitous & Necessary, but Fallible & Variable .............................. 240
    6.1.3 Ask the Experts; The Role of Theory ................................................................. 242
  6.2 NETWORK BIAS ...................................................................................................... 245
    6.2.1 Inherent Bias ..................................................................................................... 246
    6.2.2. Externally-Provoked Bias ................................................................................ 248
  6.3 VARIANCE OF EPISTEMIC INTUITIONS: WN&S ................................................... 250
    6.3.1 The Variance as Network Bias ......................................................................... 250
    6.3.2 Epistemic Variation – An Analysis ................................................................... 254
  6.4 INTUITION AND SPM ........................................................................................... 258

BIBLIOGRAPHY ............................................................................................................... 261
OVERVIEW

Intuitions are often treated as a kind of epistemic bedrock, the point at which shovels of philosophical analysis are turned. In particular, appeals to intuition are regarded as constituting the ultimate evidential source as to whether or not some theory or conceptual analysis is warranted. Intuitive judgments are solicited about imagined cases to which the theory or analysis applies; conforming to these deep-seated intuitive judgments is regarded a virtue, not conforming a vice. It is the intuitions of the folk, that is, intuitions that we all share, to which these judgments are normally deferred.

This type of analysis is pervasive enough to be dubbed “standard philosophical methodology (SPM)” by Goldman and Pust. The SPM with which I am particularly concerned is that which offers a conceptual analysis that is normative in character; that is, the resulting analysis pronounces upon the proper use or applicability of that term. It is that definition which is subjected to intuitive analysis. An instance of SPM is the analysis of ‘knowledge’ into ‘justified true belief’ - the refutation of which is meant to be supported by the intuitions generated by Gettier-type cases. The use of ‘Gettier-intuitions’ has recently come under attack by Weinberg, Nichols & Stich (WN&S) as they present evidence that the generated intuitions lack substantive normativity since they can be shown to vary from one social group to another.

I take it that the problem is quite general, however, and not limited to Gettier-cases. The folk are asked to respond to any number of imaginary thought experiments; to pronounce on some characterization of the mental life of Davidson’s Swampman – a character rising de novo with no causal history with the world yet with all the beliefs and intentional states of the reader; about the semantic status of persons in cases of Twin Earths – in which phenomenology remains the same, but the chemical makeup of the perceived properties is altered; and about any number of other thought experiments designed to push or pull a conceptual analysis in one way or another, or to inspire cognitive resonance/dissonance toward some theory of time, space or mind. The folk, then, are asked to respond intuitively to any number of bizarre scenarios, and those intuitions are taken to be evidence in some philosophical discourse.
However, it would appear *prima facie* from WN&S’s work, that the folk intuitions to which SPM refers *tout court* as evidence are simply not a reliable evidentiary source. Demonstrated variance in intuitive responses among social groups indicates that a bedrock of traditional philosophical analysis may be in danger of turning to sand. I will argue that this is the case. SPM is in error in using unfiltered intuitions of the folk as evidence.

My strategy is to provide a much-needed analysis of intuition itself, with particular emphasis on its underlying cognitive basis and resulting epistemic status. Although the philosophical literature abounds with appeals to intuition, there is a conspicuous lack of analysis of intuition itself, as a cognitive process.

A working definition of intuition is derived in Chapter 1, and the motivation behind that derivation shall be such that the definition is (as) free (as possible) from any prior commitment to any specific theory of concepts. Intuition is to be regarded as a kind of spontaneous judgment of fact, a recognition of a kind of pattern. It is, in Pascal’s words, an ability to “see the matter at once, at one glance, and not by a process of reasoning”\(^4\).

The next three chapters are a sustained attack on the traditional computational view of mind as Turing machine. This attack is to motivate the acceptance of, and initially characterize, an alternative account.

In Chapter 2 the metaphysics of patterns is investigated. The patterns themselves are identified as referents of our concepts, and I argue that most patterns of any interest are too complex to be identified by traditional Turing computation. I develop a hierarchy of complexity composed of a number of axes, corresponding to degrees of mind-dependence, context-dependence and amenability to algorithmic description. It is the last that is most relevant to my argument against mind as Turing machine. I conclude that the only tenable definition of pattern is an operational one, in which ‘recognition’ remains an essential term, but I also argue that the somewhat circular nature of this definition is not vicious: patterns are what is taken as patterns by a pattern-recognizing creature, but that creature is embedded within a socio-linguistic community, and it is that community that provides for a degree of normativity in acts of recognition.

In Chapter 3, the ‘anti-computational’ view is extended by considerations of the epistemic nature

---

3 Goldman & Pust, 2002, pg. 76.
4 Pascal, Pensees, as quoted in the Introduction to Dreyfus, 1999
of pattern recognition. ‘Recognition’ of patterns is likened to a skill, one that invokes a distinct sort of computational primitive – that of similarity-for-free - in which two patterns may be judged as relevantly similar, without the logical presupposition of identity conditions. I claim, but at this point do not argue, that only a connectionist view of mind could account for such a skill.

In each of these chapters, it is argued that it is logically impossible for traditional computing systems to identify the patterns of interest: no computing mechanism, regardless of its size or the length of time available to run the applicable algorithms, can perform the required computations. Two main lines of argument are given to this end. First, the individual elements, of which the larger pattern is to be composed and over which computations were to take place, cannot be independently identified. Second, the patterns themselves are not well-defined, and it is only a situated creature with the appropriate attachments to the world (physical, historical, emotional, cultural) that could provide the necessary disambiguation.

The stronger logical impossibility of computation need not stand for these chapters to do their work in establishing the need for an alternative computational paradigm. A mere pragmatic impossibility is sufficient. If the computing required for pattern identification can be shown to be sufficiently complex that physical limitations (size of computing device, time required of computation) indicate our brains cannot be recognizing the patterns of interest by making use of the brute force of traditional computing methods, then the argument stands.

In Chapter 4 independent considerations, over and above those related to the patterns themselves, are provided for rejecting Fodor’s Language of Thought theory of mind. I reject what I call the psychological desideratum - that which is meant to be explained by the LOT thesis – and so reject that which motivates the view. The psychological desiderata are a cluster of claims that centre around the assertion that our thinking is perfectly systematic, productive, semantically discrete; our thinking is ‘digital’ in nature. Fodor argues that the only way to account for such structure is to posit that it is reflective of ‘deeper’ computational structure. I argue Fodor is mistaken in characterizing our thought in this manner. I do find a sympathetic reading of Fodor’s atomistic view of concepts, however, one that is amenable to a connectionist interpretation.

In Chapter 5 I develop my alternative account, and provide an analysis of intuition and theory of concepts which I call ‘pragmatic conceptualism’. The majority of the original work in this thesis is contained in this chapter. I mean to provoke a commitment to taking the implementation story seriously, and a commensurate acceptance of the repercussions that follow. Intuitive judgments and
concepts are each analyzed into an appropriately defined set of signals passing through the network, and have two components: a causally efficacious mental component and a semantic public component. Consciousness is given a pivotal role in linking the public semantic content of a concept or intuition to the causally efficacious signal associated with that concept or intuition. Intuitions are shown to be instantiations of concepts, they are therefore, direct evidence of a person’s concepts. The set of weights and connections that comprise a network is identified as a kind of knowledge, and is highly dependent on the unique history of interactions that network has had with the world. Both concepts and intuitions reflect that unique knowledge, and as such, are expected to vary from person to person.

In Chapter 6, I return to the use of intuition in SPM, and Gettier cases in particular, with a view to using my analysis of intuition to explain the WN&S data as a kind of ‘network bias’. Intuitions will vary among groups (indeed it would be surprising if they did not) and so if they are to be used as evidence in philosophical analysis, need to be incorporated into a larger methodology which can account for, and accommodate, that variance.

Methodological Considerations

Underpinning (and resulting from) work in this thesis is a commitment to two variations on a general theme, that I call epistemic and ontological continuity. These views are simply asserted at this early point for clarity, but they are defended as reasonable at various points throughout the thesis. I take support from Fodor; “it’s the usual ... assumption: if my story is plausible, that argues for my story. For the moment, all I ask is the temporary suspension of your disbelief”.

There is an important methodological consideration resulting from these commitments.

Epistemic continuity states that conception and perception are points on a continuum; that is, although conception can be identified as distinct from perception in ways that are useful (just as a pebble can be identified as distinct from a boulder), the difference masks a large degree of interpenetration and overlap. Perception of everyday objects involves concepts in order to see the object as an object - we perceive meaningfully - and concepts are not (generally) isolable from the acts of perceiving that generate that concept. I betray an empiricist leaning here, in that I hold that

---

5 Denoted as concept(X) _mental_ or intuition(X) _mental_ if the semantic target is X; e.g concept(KNOWLEDGE) _mental_ is a signal corresponding to a thought about knowledge.

6 Denoted as concept(X) _public_ or intuition(X) _public_ if the semantic target is X; e.g. concept(KNOWLEDGE) _public_ is taken to be what we, as a socio-linguistic community, mean by the word ‘knowledge’.

7 Fodor, 1998, pg. 73
concepts are derived (in large part) from acts of perception. The ability to generate intuitions about concepts and conceptual analysis is continuous, as a cognitive skill, with the ability to generate intuitions about perception and perceived objects or states-of-affairs.

In the earlier quote from Pascal, “We must see the matter at once, at one glance, ...”, the word ‘matter’ is now interpreted quite liberally, and under these considerations of continuity, it could refer to some state of affairs that sits before one (like a chess game, with the ‘matter’ being a weak king side) or some imagined case upon which one must pass judgment (like a Gettier case). One perceives the matter at once, even though the matter can be described in a way that is conceptual in nature (the imagined case of semantic assent of abstract meaning, like ‘knowledge’), or perceptual (the apple about to fall).

Ontological continuity states that the physical patterns normally associated with perceived objects (such as a tree or cup) are contained on a continuum with the more abstract patterns associated with semantic practice (such as identifying an act as ‘graceful’, ‘relevant’, or as an instance of ‘knowledge’). Here I side with Davidson, whose notion of an ‘event’ can be distributed quite arbitrarily throughout space and time, and can be captured in natural language – even if the event occurs only once. A ‘graceful’ act may be arbitrarily complex in terms of the identification conditions over space and time, but the event can be characterized by a single word ‘graceful’, or even captured in its entirely with a proposition; ‘The graceful jump by Rolling Thunder at the National Equestrian Championships’.

The methodological consideration regarding these two sorts of continua - an epistemic continuum with perception at one end and conception at the other, and an ontological continuum with simple perceptive patterns at one end and the complex conditions normally associated with intentional items - is that the principles gleaned from one end of the continuum can be applied, in a stronger sense than merely analogical, to the other. Although I hope that this methodological consideration shall be clearly authorized by the time the work of Chapters 2 and 3 on ‘patterns’ and ‘recognition’ is done, it should be noted early on, as the treatment of matters conceptual and perceptual are taken as relevantly similar from the start.

I shall use the terms ‘metaphysics’ and ‘ontology’ somewhat interchangeably since they are related – following the Oxford Dictionary of Philosophy, ontology is a “branch of metaphysics that concerns itself with what exists”8 – although my use of ‘metaphysics’ emphasizes an irreducibly

8 Blackburn, 1996, pg. 269
conceptual component of how we categorize that which exists.

ARGUMENT FORM

The overall argument form is as follows:

1) Intuition is a spontaneous judgment of fact that can be likened to a sort of pattern recognition. (Chapter 1)

2) That pattern recognizing behaviour cannot be accounted for by traditional, Turing computing. Three arguments are provided to this end:
   a. The patterns themselves are too complex (Chapter 2).
   b. The means by which humans perform pattern-recognizing tasks is not by leveraging traditional computational resources (Chapter 3).
   c. The strongest expression of the Turing model, Fodor’s Language of Thought view of mind, is denied for independent considerations. (Chapter 4).

3) Pattern-recognizing behaviour is a product of our connectionist networks, and intuition is contingently shaped by those networks (Chapter 5).
   a. It is an empirical fact that connectionist networks do exhibit the relevant pattern-recognizing behaviour.
   b. People’s brains share basic architecture with connectionist networks.
   c. Hence, the reason we are capable of recognizing patterns is because we have access to connectionist network computing – Argument by abduction.
   d. Our cognition is driven by a connectionist network, and there are ramifications for the way in which we think generally, w.r.t. intuition and concepts in particular – Corollary: “Taking Connectionism Seriously.”
   e. Intuition is a particular sort of signal within a connectionist network, and so is highly contingent on the training profile of that network.

4) Intuition, if used in the context of philosophical discourse, is expected to vary between cultures. The proper role for intuition is within a larger methodology that filters for this variance. Philosophy, in taking folk intuitions as a reliable evidential source, is committing a gross error.
Chapter One

1. INTUITION

1.1 Introduction

Analytic philosophy is riddled with references to intuition, but very little attention has been paid as to what those intuitions are or why one is justified in appealing to them. Intuitions are often appealed to as the final arbiter in what would otherwise appear to be an argumentative impasse, or as a guide to further analysis when philosophical shovels hit bedrock, yet the reasons why such foundational status should be granted are generally not given. Presumably, it is intuitively obvious.

Williamson seeks (in vain) for an account of intuition that might explain how it is that our intuitions are meant to be indicators of truth; “... analytic philosophy has no agreed or even popular account of how intuition might work, no accepted explanation of the hoped-for correlation between our having an intuition that P and its being true that P.”9 With a similar sense of scepticism, Hintikka asks, “[F]or what is supposed to be the justification of such appeals to intuition? One searches the literature in vain for a serious attempt to provide such a justification”10. I leave it to the reader to thumb through the indexes of any number of philosophical texts to satisfy herself of the truth of this statement. This thesis is an attempt to provide such an analysis.

Having one’s conceptual analysis – of, say, ‘personal identity’, ‘causation’ or ‘knowledge’ - satisfy appeals to intuition is a virtue, failing to satisfy them is a vice. What I will call (from Goldman & Pust) Standard Philosophical Methodology (SPM), attempts to provide a definition or analysis11 of a concept, and does so by explicit appeals to intuition. Those appeals are couched in the form of judgments as to whether some imagined case is (or is not) an instance of the concept or predicate under scrutiny. The cases are constructed so that the intuitions are meant to provide counter-evidence to the adequacy of the analysis. In the normal course of events, the case is meant indicate that a definition does not fulfil the sufficiency requirement if the case passes intuitional scrutiny as an instance, but does not meet the definition. Likewise, if the case does not pass intuitional scrutiny as an instance, but does meet the definition, then the definition is deemed not to

---

9 Williamson, Philosophical Intuitions, pg .109
10 Hintikka, 1999, pg. 130
11 By way of a set of necessary and sufficient conditions or some other well-constructed set of constraints to flesh out the concept, be they possession conditions, the specification conditions of the proper causal relationship for a causal theory of reference, or some other sophisticated account. These details can be left aside for now, but I will commit SPM to a definitionist account.
fulfil the *necessity* requirement. There are variations on this theme, but such philosophical activity is typical.

Goldman & Pust see this activity as a two-state process. In the first stage, an intuition that \( p \) is “taken as (prima facie) evidence of the truth of \( p \)”\(^1\). In the second stage “the truth of \( p \) is used as positive or negative evidence for the truth of a general theory”\(^1\). I am here concerned with the validity of the first stage of the process, whether or not the intuitions should be assumed to attain a high enough epistemic status that they *should* be used as evidence in the second stage.

The persons to whom one appeals for such intuitions on the subject are the folk. Ordinary folk, and not experts laden with theoretical biases, are meant to be the clearest expression of these intuitions; “philosophers rightly prefer informants who can provide *pre-theoretical* intuitions about the targets of philosophical analysis, rather than informants who have a theoretical ‘stake’ or ‘axe to grind’”\(^1\). While contamination of judgment by prior theoretical commitment may indeed be something to avoid, what I aim to show is that reliance on the uninformed intuition of the folk poses no less a danger to philosophical analysis. One wants, so to speak, to sail the epistemic ship between the Scylla of the well-informed but over-committed opinions of the expert and the Charybdis of the uninformed naivety of the folk.

A typical example of SPM is the set of Gettier cases taken as counter-examples to a justified-true-belief (JTB) conception of knowledge. The imagined case is one in which a person has a justified true belief, but one nevertheless intuitively withholds assent to attribution of knowledge. There are also other types of epistemological scenarios that are designed to elicit intuitions about knowledge. For example, in many, there remains some ‘accidental’ character to the knowledge, or some distinction that the presenter of the case is trying to bring to bear, such as the need for justification to be internally accessible; “An issue of great moment in recent analytic epistemology is the internalism/externalism debate. ... Components of an agent’s doxastic situation available to introspection are internalistically kosher; other factors beyond the scope of introspection, such as the reliability of the psychological mechanisms that actually produced the belief, are epistemically external to the agent.”\(^1\)

Here, one constructs a case of reliable belief acquisition without internal justification and the case is meant to elicit intuitive judgments that withhold ascriptions of

---

\(^1\) Goldman & Pust, 2002, pg. 77
\(^1\) Goldman & Pust, 2002, pg. 77
\(^1\) Goldman & Pust, 2002, pg. 77
\(^1\) Goldman & Pust, 2002, pg. 78
\(^1\) Weinberg, Nichols and Stich, 2001, pg. 15
knowledge. A typical\textsuperscript{16} epistemological scenario as presented by Weinberg, Nichols and Stich (WN&S) is the following:

One day Charles is suddenly knocked out by a falling rock, and his brain becomes re-wired so that he is always absolutely right whenever he estimates the temperature where he is. Charles is completely unaware that his brain has been altered in this way. A few weeks later, this brain re-wiring leads him to believe that it is 71 degrees in his room. Apart from his estimation, he has no other reasons to think that it is 71 degrees. In fact, it is at that time 71 degrees in his room. Does Charles really know that it was 71 degrees in the room, or does he only believe it?\textsuperscript{17}

The intuitive withholding of assent is meant to indicate that the conception as analyzed is mistaken, or incomplete – in this case, Charles does not have epistemic access to the reasons why his belief is justified. The intuitive judgment is to count as evidence against an externalist theory of knowledge; indeed, such evidence is taken to be so strong as to refute the conception.

Recently, doubts have been raised as to the legitimacy of appeals to intuition, particularly in these Gettier-type cases. WN&S have found evidence that the ordinary folks’ intuitions do not exhibit the requisite normativity required by a SPM analysis of knowledge. They have found evidence that different sub-cultures of the ordinary folk exhibit different intuitions for the same imagined Gettier cases. If such intuitions cannot be counted upon to exhibit stability across cultures, and those intuitions are appealed to as final arbiter in a conceptual analysis, whence will come the desired normative constraints on the concept itself? Under SPM, knowledge is not the sort of concept that one wants to abandon to the whims of cultural relativity, it is meant to support a deeper, more stable sort of conceptual analysis. I will address in detail the issue of conceptual normativity, as a virtue and a vice, but for now, it is sufficient to say that differing intuitions across cultures ignite a kind of philosophical crisis. Prima facie, for an epistemologist, perhaps the resulting casualty should be the elevated status of intuition.

The problem pointed out by WN&S is a specific example of a more general problem. The intuitions of the folk have often been shown to be at odds with more considered opinion;

\textsuperscript{16} The literature is filled with myriad examples of Gettier cases and other epistemological scenarios, and it matters little at this point which particular case I emphasize. To establish my point I only need to get across the typical format, presentation and resultant status of the elicited intuitions. I shall freely call all such cases Gettier cases, although strictly speaking, there was an original case presented by Gettier, and a raft of Gettier-type cases that followed.
“Twentieth-century physics, for example, has notoriously shown the inaccuracy of our everyday intuitions as hypotheses about the nature of space and time, light, and the microscopic world. The well-cultivated intuitions of a theoretical physicist may be more accurate in this regard, to the extent her physical theories work their way underground into her intuitive judgments.”\(^{18}\). Does this mean that we want to claim that such folksy intuitions are simply wrong? Surely, it’s not that simple: “In contrast, our commonsense intuitions about the behaviour of mid-sized objects, as we interact with them in everyday life, are generally rather good ...”\(^{19}\).

As I see it there are (at least) three relevant questions with regard to intuition. The first question relates to the experience of intuition as a psychological phenomenon - how is intuition distinct from other sorts of mental activity? The second question relates to the cognitive mechanism that may underlie that psychological phenomenon - why is intuition different from other mental phenomena, or what explains that difference? Finally, what are the epistemic consequences of that distinction - should intuition be considered a valid evidential source? The purpose of this chapter is, in part, to answer the first question. I will develop a definition of intuition, as a distinct mental phenomenon, that is independent of any prior commitment to a theory of concepts or meaning. With regard to the second question, an account of the cognitive mechanism that gives rise to the phenomenological and intentional character of intuition will be developed in Chapter 5. The last question, of over-arching concern, is answered in the final chapter.

In this chapter then, will be the following. First, an outline of the ‘normativity problem’ is given in Intuitions About Knowledge Ascriptions: Weinberg, Nichols and Stich’s Normativity Problem (1.2). I shall then follow Gopnik and Schwitzgebel in disambiguating between two attitudes one can adopt toward intuition in Folk Intuitions: Evidence of What? Hypotheses vs. Evidence (1.3). It is contended that SPM is often guilty of conflating these two attitudes. I will also point out some evidence that our intuitions, when regarded as hypotheses about the world, can be outright wrong. I then give Two Contemporary Analyses of Intuition; Goldman & Pust, Pust (1.4). Following from these analyses I shall develop my own definition of intuition in Intuition, A Definition (1.5), and make some Concluding Remarks (1.6).

\(^{17}\) WN&S, pg. 15
\(^{18}\) Gopnik & Schwitzgebel, pg. 78
\(^{19}\) Gopnik & Schwitzgebel, pg. 78
1.2 Intuitions About Knowledge Ascriptions: Weinberg, Nichols and Stich’s Normativity Problem

1.2.1 The Normative and Descriptive Projects

Weinberg, Nichols and Stich (WN&S) differentiate between four projects associated with SPM; the Normative, the Descriptive, the Evaluative and the Ameliorative Projects. The Normative project is “the most philosophically central of the four” and it “attempts to establish norms to guide our epistemic efforts.” The explicitly normative character of the Normative Project is that its role would be to endorse some methodology as the correct (or incorrect) way of going about our epistemic affairs. The implicitly normative aspect of SPM itself then, is that it - as a chosen philosophical methodology - endorses a specific method of going about a particular quest for knowledge. The role of the Normative Project is to make explicit just that issue; its role is in “specifying which ways of going about the quest for knowledge should be pursued and which should not.”

The Descriptive Project takes on the task of describing - or analyzing - concepts or language. If concepts are to be described, then the target is “the epistemic concepts that some group of people actually invoke.” If it is language that is being described, then the target is “the way some group of people use [epistemic] language or to analyze the meaning of their [epistemic] terms.” The point here is that the Descriptive Project takes what are existent concepts or predicates, used or employed in some way by a group of people (or tellingly, by what is often presumed to be ‘people’ qua people, ie, all people in general) and describes or analyzes what those concepts are, or what those predicates mean. SPM is a version of the Descriptive Project with a method that often employs intuitions about imaginary cases to confirm or deny the correctness of some theory, concept or predicate.

The two projects are related. The Descriptive Project generates what it is that we mean by ‘knowledge’. ‘Knowledge’ itself is implicitly normative, as it is the epistemic good that the

---

20 Weinberg, Nichols and Stich, 2001, pg. 2
21 Weinberg, Nichols and Stich, 2001, pg. 2
22 Weinberg, Nichols and Stich, 2001, pg. 2
23 Weinberg, Nichols and Stich, 2001, pg. 2
24 Weinberg, Nichols and Stich, 2001, pg. 3
Normative Project is after. So the Descriptive Project is generating the meaning of the term that defines the epistemic good of the Normative Project.

These two projects, the Normative and the Descriptive, are of central importance - however, for completeness: the Evaluative Project “tries to assess how well or poorly people’s actual belief forming practices accord with the norms specified in the Normative Project” and the Ameliorative Project attempts to address any deficit found by the Evaluative Project. On this model, my analysis of intuition is an instance of the Evaluative Project - in that I subject the epistemic status of intuition to scrutiny - and my recommended treatment of intuition is itself an Ameliorative Project.

Note that the Descriptive Project by itself does not necessarily contain any presumptions about what concepts or predicates belong to what groups of people, or even what concepts or predicates actually are. It does not, by itself, presuppose a theory of concepts, it only purports to go about the project of describing concepts as they are actually used by some group of people. One could be investigating the concept ‘space’ as it obtains amongst physicists at Princeton, or as it obtains among the Aboriginals of Papua New Guinea. One could assume that ‘space’ must be a concept such that it must be the same over those groups of people, or one could assume that it varies from person to person. It is only in implementing a particular version of a Descriptive Theory that one ‘shows one’s hand’, so to speak. SPM, as I’ve articulated it, is not neutral with respect to these issues.

1.2.2 Epistemic Romanticism & Intuition Driven Romanticism

WN&S uses the term Epistemic Romanticism to capture the idea that “knowledge of the correct epistemic norms (or information that can lead to knowledge of the correct norms) is implanted within us in some way, and with the proper process of self-exploration we can discover them.”

Epistemic Romanticism is a sub-category of a more general 19th century Romanticism, which WN&S characterize as the idea that “our real selves, the essence of our identity, is implanted within us, and that to discover who we really are we need but let that real identify emerge.”

SPM is not just committed to the idea that whatever ‘stuff’ is needed to flesh out the concept

---

25 Weinberg, Nichols and Stich, 2001, pg. 3
26 Granted that intuitions are distinct from beliefs, as I shall make clear presently.
27 Weinberg, Nichols and Stich, 2001, pg. 4
'knowledge' is there, sitting in the heads of the folk, but it is using - to borrow a term from Dennett - intuition pumps to get at that stuff. A Gettier case is presented, the folk sit and think, and voila!, they pronounce that their deepest intuition about the matter is that such-and-such is definitely not (or definitely is) an instance of knowledge. The explicit appeal to intuition, as a means to get at the stuff in the heads of the folk that will flesh out the concept ‘knowledge’, WN&S call Intuition Driven Romanticism (IDR). Of the Gettier cases that of particular concern here;

... we think a plausible case can be made that a fair amount of what goes on in normative epistemology can be classified as Intuition Driven Romanticism. Moreover, to the extent that it is assumed to have normative implications, much of what has been written in descriptive epistemology in recent decades also counts as Intuition Driven Romanticism. For example, just about all of the vast literature that arose in response to Gettier’s classic paper uses intuitions about specific cases to test proposed analyses of the concept of knowledge.

1.2.3 The Normativity Problem

So what, exactly, is the problem? Perhaps the concept knowledge can be fleshed out in just such a way. The problem is that there is empirical evidence that intuitions about the applicability of definitions of knowledge - as is the case for many intuitions of the folk - vary from group to group, and indeed, from time to time for a single individual. WN&S provide evidence that there is significant variance in intuitive responses to various Gettier cases across cultures, and across individuals over time (depending on the order in which cases are presented to them). Given some facts about cultural differences in cognition, one can jig Gettier cases to exploit those differences, leading to large discrepancies in the statistics of judgments made about those sorts of cases. The evidence to back this claim shall be detailed in the final chapter once both ‘intuition’ and ‘concept’ have been given some analysis. For now, I simply assert that there is a problem: intuitions regarding Gettier cases vary culturally, and temporally over a single individual.

28 Weinberg, Nichols and Stich, 2001, pg. 4
29 Weinberg, Nichols and Stich, 2001, pg. 8
30 That there are such differences is claimed by Nisbett and Haidt. The details of this claim will be made clear in Chapter 5.
Well, is that a problem? Perhaps one can accept a kind of relativisation of the concept ‘knowledge’, from culture to culture and from time to time for an individual. Perhaps, and a variant of that view is one I shall defend – the degree of variance, and how that variance is minimized, though, are key components of a much larger methodological framework by which intuitions may be accommodated. That story will come out in the final chapter. However, this is not an option for SPM. SPM has an explicit commitment to universally applicable definitions, and to intuition as a form of evidence in developing those definitions. The SPM analysis of knowledge, is a form of IDR and “yield(s) as outputs claims that putatively have normative force.”

What one is doing in developing a JTB analysis of knowledge in the first place was meant to be - as is meant to be the case in much, if not all, of SPM - generating necessary and sufficient conditions for the application of the concept knowledge. Knowledge is not, on this view, relative to a culture, nor is it meant to vary over time for an individual; knowledge is knowledge is knowledge. If intuitions vary, and these are used as data in SPM, then “it looks like the IDR strategy for answering normative epistemic questions might sanction any of a wide variety of regulative and valuational norms. And that sounds like bad news for an advocate of the IDR strategy, since the strategy doesn’t tell us what we really want to know. It doesn’t tell us how we should go about the business of forming and revising our beliefs.”

One can push the problem back to the question of selecting a privileged set of intuitions, but “it is less than obvious (to put it mildly) how this move could be defended. Why should we privilege our intuitions rather than the intuitions of some other group?”

What’s at stake then, is whether or not it is tenable to hang onto both of the commitments I have attributed to SPM. Is it possible to be committed to conceptual definitionism (or some close variant of conceptual theory) and sanction appeals to intuition in developing those definitions? This problem can be put in the form of a trilemma. One could a) abandon folk intuition as a form of evidence, b) deny that intuitions vary from group to group, c) deny that the concepts behind predicates like ‘knowledge’ are universal, that they have cross-cultural normative force.

Abandoning intuition - option a) - is possible, but there are huge swaths of philosophy one would also have to abandon. The basis of SPM would be shaken to the core. A variant of a) is to deny folk

32 Weinberg, Nichols and Stich, 2001, pg. 9
33 This is a question I shall answer, as there are ways of privileging one set of intuitions over another once one has given a cognitive explanation for the content of intuitive judgments. That story must wait.
34 Weinberg, Nichols and Stich, 2001, pg. 9
intuitions but to allow expert intuitions. This is a position I shall endorse. Denying that intuitions vary - option b) - is not a live option. It seems an empirical fact that they do vary, given both WN&S’s evidence, as well as that which will be presented in the body of this thesis. Option c), adopting a different theory of concepts, in which it is possible for conceptual content to vary across groups or individuals, is also one that I shall endorse. Call this option *epistemic relativism*. Note that a) and c) can be made to be compatible; expert opinion can be solicited in order to *reduce* but not eliminate variance in the accepted meaning of a concept.

Thus, my position will be - roughly - that intuitions are fine as evidence as long as they are the intuitions of people who have the requisite experience with the phenomenon in question. A theory of concepts that both accommodates and explains the variance in intuition is pragmatic conceptualism (although this position will be separately motivated by matters metaphysical and epistemic). There may always remain *some* variance in the definitions endorsed by the intuitions of experts, but such variance may be minimized by ongoing mutual discussion and influence. That is, the ‘normativity problem’ is to be solved by the very pragmatic notion of individual development of familiarity with the subject matter, followed by public discussion and attempts at inter-subjective agreement.

---

35 Or indeed, cross-species, since aliens may also be the legitimate target of an analysis of ‘knowledge’.
1.3 Folk Intuitions; Evidence of What? Hypotheses vs. Evidence

As a first-order refinement of intuition, I take note that Gopnik & Schwitzgebel (G&S) point out an important distinction between taking intuitions to be psychological data in need of explanation and taking them to be hypotheses about the world. It is my view that the intuition of the folk are to be taken as psychological data, and they may or may not be legitimated as hypotheses about the world.

In the latter case, “we use our intuitions as a source of empirical hypotheses. We might think of intuitive judgments as particularly plausible initial hypotheses about the nature of the world.” It is as a treatment of intuitions as hypotheses about the world that Gettier cases fall. It is a hypothesis about knowledge itself that we take such intuitions to be testing. For G&S such intuitions can provide a starting point for further intellectual investigation, indeed, are often the only way to get any such an investigation off the ground; “... it is probably impossible to start a science from scratch with hypotheses and assumptions that are entirely based on observation or experiment. One must begin with intuition and correct it with experiment.” Such intuitions are, for G&S, like any other hypotheses and as such are defeasible, and vary depending on the experience and expertise of the intuiter; “Like all such hypotheses, our intuitions are accurate as a guide to the world to different degrees, depending on both the person who has those intuitions and the subject area to which those intuitions apply ...Our intuitions can develop substantially as our expertise grows.”

I am entirely sympathetic with this description of intuition, and during the course of this thesis, will fill out an account that agrees to a large extent with this position. I take particular note of their characterization of intuition as being a skill that can develop over time, becoming sharper and more accurate, and perhaps being able to differentiate ever-finer conceptual distinctions. For G&S, there is a direct link between training and quality of intuitive judgments; “The well-cultivated intuitions of a theoretical physicist may be more accurate [with regard to hypotheses about space, time and light], to the extent her physical theories work their way underground into her intuitive judgments.”

Notoriously, our initial hypotheses can be plain wrong, and wrong not about a matter in which the correct hypotheses would be some complicated affair, derived after considered intellectual effort,

---

36 Gopnik & Schwitzgebel, 2002, pg. 78
37 Gopnik & Schwitzgebel, 2002, pg. 79
38 Gopnik & Schwitzgebel, 2002, pp 78-79
such as that which can explain the nature of space and time. We can be wrong about predictions relating to our own phenomenological states. Wisniewski provides a number of examples. In one such example, one would normally intuit that people would prefer less pain over more pain – clearly a folk intuition if ever there was one. Surprisingly, this can be shown to be false. Our judgment of pain experience can be demonstrated to be dominated by the last few moments, leading to quite counter-intuitive results. Wisniewski cites an experiment\textsuperscript{40} in which people were exposed to 60 seconds of cold water, then an experience of an identical 60 seconds followed by 30 more seconds in which the water is gradually warmed. Sixty-nine percent of participants indicated they preferred the second experience, thus choosing more pain - our initial intuitions, as a hypothesis about pain preference, are confounded.

Once we learn about our predilection to place a larger weighting on the final moments of the experience, however, our intuitions regarding the case would probably change. Indeed, if we went through the experience itself, or some similar experience, it seems reasonable to assume that our initial hypotheses about pain tolerance would become more nuanced, and take into account the temporal nature of the experience about which we are to make an intuitive judgment. Here, it seems, “people never even consider the right intuitions [or hypotheses] in the first place. The intuitions are not considered because they involve influences of which people are not consciously aware. For example, when people make judgments about pain they may simply be unaware that sometimes they are affected by the peak and final moments of pain but not affected by its duration”\textsuperscript{41}

Intuitions as hypotheses about the world then, can be quite wrong, but can develop over time. This should not be surprising.

G&S go on to describe what it means to treat intuitions as data, as a psychological state, itself in need of explanation. Here, the evidence is provided, not as a hypothesis about the world (nor about the acceptability of some hypothesis about the world), but about the state of mind of the individual intuiter (or of mind writ large, if the intuition is general enough to be similar across individuals). Here, “we think of intuitions as data that need to be explained by a psychological theory.”\textsuperscript{42} Such intuitions are not taken to be an accurate account of the phenomena they purport to explain, but are taken to be accurate reports of a psychological state that, in turn, stands in need of explanation; “A psychological theory has to explain why we have these intuitive beliefs even if, indeed especially if,

\textsuperscript{39} Gopnik & Schwitzgебel, 2002. pg. 78
\textsuperscript{40} Kahneman, Fredrickson, Schrieber and Redelineir, 1993.
\textsuperscript{41} Wisniewski, 2002, pg. 50
the beliefs are quite false ... Intuitive judgments are an important part of our mental lives; psychological theory, therefore, aims to explain the nature and origin of those intuitions.”

Although G&S equivocate between ‘belief’ and ‘judgment’, I take their point to be entirely sympathetic to my current project of providing for a cognitive explanation of intuition, prior to passing judgment on the epistemic status of those intuitions.

The problem, as far as there appears to be a need to reconcile the differences between folk intuitions and expert intuition, or between inter-cultural groups passing judgment on a Gettier case, is that these two views of intuition are often conflated. Folk intuitions are indeed accurate reflections of what the folk think, and as such, are perfectly good data to use in assessing or developing a psychological theory. However, that does not mean that they are accurate reflections of the hypothesis upon which they pass judgment. The conflation is probably most prevalent in the philosophy of mind, since it is often the folk psychological states themselves which are under scrutiny. The folks’ theory of mind, for example, is expressed by their intuitions, and as data about the folks’ theory of mind is logically unassailable. The intuitions are taken to be true, as they reflect their own subject matter. This does not mean, however, that the folks’ intuitions “about the mind, treated as hypotheses about human cognition, are similarly untouchable by empirical observation.” Perhaps the confusion can be excused in the philosophy of mind, but it does appear to be a general problem. For G&S, this is a “category mistake”;

“Within philosophy of mind, there has been a tradition of either accidentally confusing different uses of intuition, or deliberately treating them as identical, and consequently drawing inappropriate conclusions. An unspoken assumption of much argumentation in the philosophy of mind has been that to articulate our folk psychological intuitions, our ordinary conceptions of belief, truth, meaning, and so forth, is itself sufficient to give a theoretical account of what belief, truth, meaning and so forth, actually are.”

That IDR purports to generate intuitions from the folk, and have those intuitions form the grounds for acceptance or dismissal of a theory of knowledge, places IDR squarely in the G&S’s sights on this charge.

42 Gopnik & Schwitzgebel, 2002, pg. 79
43 Gopnik & Schwitzgebel, 2002, pg. 80
44 Gopnik & Schwitzgebel, 2002, pg. 85
45 Gopnik & Schwitzgebel, 2002, pg. 84
46 Gopnik & Schwitzgebel, 2002, pg. 84
The larger question of the status of the relationship between folk theories (or folk intuitions) and scientific theories (or expert intuitions) that purport to provide for a more sophisticated analysis of that phenomena is a critical one, and answering it by simply asserting that the scientific theory replaces the folk theory does not do justice to the pragmatic usefulness of folk theories and the accompanying ontology that furnishes our everyday lives. Churchland, notoriously, eschews talk of beliefs and desires, as the more sophisticated neurology that is to replace folk psychology finds no such items in the brain. Are we to abandon talk of beliefs and desires? Surely, the folk intuitions in this case provide for some kind of hypothesis about the world, even if the neurologists find no correlates.

The status of folk intuitions as comprising accurate data about how the folk think, however, is clearly not to be confused with the status of folk intuitions as hypotheses of the folk about some phenomena that may or may not be better explained by experts. The question that must be answered then, in treating folk intuitions as hypotheses, is: does a more sophisticated analysis (which may be quite counter-intuitive) of the phenomena in question better explain the phenomena, or explain it away? In either case, the folk intuitions are to be treated as data in a psychological explanation. If the phenomenon is successfully explained away - as perhaps ghosts are and beliefs and desires are not - then one will want to simply say the folk intuitions are wrong as hypotheses about the world. If the phenomenon is better explained, then one will want to say that the folk intuitions are a kind of initial hypothesis, a good rough-and-ready start to the analysis. In either case, the initial intuitions stand in need of explanation as to why they took the form they did. Folk physics is surely not wrong as I walk across the street avoiding cars, it is

---

47 Or has not yet found such co-relates. In my view, to think such items will be found in the brain is another kind of category mistake. See Chapter 5.
48 There is, of course, a subsidiary question, namely: what does it mean to provide a better explanation? Explanatory schemes can place different emphases on, for example, the provision of reductive analysis, predictive success, coherence with other theoretical concerns or schemes, simplicity and so on. I am able to skirt this somewhat complicated issue because, for the purpose of the Gettier cases and other philosophical analyses of which are of particular concern here, a ‘better explanation’ will depend, not surprisingly, on the goal of the analysis itself. If the goal is to provide necessary and sufficient conditions for a conceptual definition then the success of explanation of the predicate or concept in question will depend on the degree to which those criterion are met. Likewise, if the goal is to develop a pragmatically-defined set of recognitional abilities that relate to the concept, or a set of minimal conditions for meaningful discourse, then a ‘better explanation’ likewise applies to the fulfillment of either of those criteria. What exactly, in my view, those criteria should be, will be made clear in my commitment to the position I call conceptual pragmatism.
49 Successfully explaining something away I take to mean something like: there is an alternate explanation of the phenomenon that allows for a continued, but altered, encounter with the phenomenon. The phenomenon is not denied a place in the ontological furniture, but it’s character (or status) changes. A ghost, for example, becomes a ‘shadow of the mind in a highly agitated state’ or ‘a play of light on the wall’.
surely a good rough-and-ready guide to the world in which I am coping.

What I intend to do in this thesis is explicitly uncouple the two notions of intuition. My analysis of intuition is meant to take intuition as data to be given an explanation, provide that explanation, and only then judge how well the folks’ intuitions stand up as evidence for or against hypotheses about the world, particularly, for ascriptions of knowledge\textsuperscript{50}.

\textsuperscript{50} To give away the ending: they shall not fare well, unless they are being used to define and defend the folk theory itself. They are necessary for the definition of the folk theory, but are not sufficient for its defence: further criteria are required if the folk theory is to be regarded as explanatorily superior to some other, more sophisticated theory.
1.4 Two Contemporary Analyses of Intuition; Goldman & Pust, Pust

I begin the analysis of intuition, regarded as a distinct mental phenomenon, with two contemporary analyses. Although I will find both unsatisfactory, I will take from each some key characteristics, using them as a starting point in developing my own definition.

1.4.1 Goldman & Pust; The ‘Reliabilist’ View

Goldman and Pust’s (G&P) main concern regarding intuition reflects the concerns of G&S outlined above. They put their concern thus: “The question that needs to be raised, therefore, is whether intuition really is a basic evidential source, in particular, whether it satisfies [the condition of being a reliable indicator of the truth of their content]”\(^51\). By ‘basic evidential source’, G&P are capturing the idea that some kinds of evidence are reliable indicators of the truth of their contents, by virtue of the form that evidence takes (given some caveats regarding favourable conditions, and so on, which I shall ignore for now). Sensory content, for example, is considered to be a basic evidential source\(^52\). G&P are concerned with the basic intuitions at issue here - those which are used as evidence in analytic philosophy. G&P call such analysis ‘standard philosophical methodology (SPM)’ in which “the content of a typical intuition is a proposition about whether a case or example is an instance of a certain kind, concept, or predicate ... to the effect that such-and-such an example is or is not an instance of knowledge [of justice, of personal identity ...]”\(^53\). Like G&S, they are enquiring as to the epistemic status of such intuitions.

A criterion by which to credit an evidential source as being basic in this way is the ‘reliability indicatorship’, which they define as:

\[
\text{“(RI) Mental states of type M constitute a basic evidential source only if M-states are reliable indicators of the truth of their contents [M-Contents] (or the truth of closely related contents), at least when the M-states occur in M-favourable}\]

\(^51\) Goldman & Pust, Pathways, pg. 77
\(^52\) That there can be a kind of deep scepticism regarding the purported content of a sensory experience- the tree I see may not be before me, although it is true that I do see a tree before me - is the kind of deep scepticism that worries Williamson (2004) and it is not the kind of scepticism of which I am presently concerned.
\(^53\) Goldman & Pust, Pathways, pg. 76
The contents of such states must be generally true. I shall not quibble about M-favourable circumstances here. For G&P, counter-factual dependence and the existence of a causal route between M-states and M-contents, are each good indicators that (RI) obtains, but are not necessary. The lack of a causal route between the M-state and M-contents is not enough to withhold from the M-state the status of basic evidential source, but “if we believed that there is no such causal route, there would be grounds for doubting that there are counterfactual dependences of the indicated sort. And if there were no counterfactual dependencies of the indicated sort, there would be grounds for doubting that the reliable indicatorship relation obtains”\textsuperscript{55}. Lack of causation or counter-factual dependence provides reason to doubt, then, but does not provide reason enough to withhold (RI) from some M-state.

G&P then derive a conception of intuition, starting off with intuitions being “some sort of spontaneous mental judgments”\textsuperscript{56} and go on to qualify those judgments as being comprised of “singular classificational propositions”\textsuperscript{57}. They then move toward their fully-fleshed out definition of intuition by way of a denial that classificational propositions of the form ‘C is P’ can be satisfied by any ‘extra-mental’, or non-psychological, content. They are motivated, then, by a prior commitment to a philosophical view they call ‘mentalism’, a view that regards all conceptual content as “a psychological structure or state that underpins a cognizer’s deployment of a natural-language predicate”\textsuperscript{58}. I remain agnostic, for the present, on this important question\textsuperscript{59}.

However, once G&P position themselves as ‘mentalists’, such that the content that intuition is evidence for is itself psychological - a definition of ‘mentalist’ they have explicitly endorsed - they

\begin{thebibliography}{9}
\bibitem{GoldmanPust76} Goldman & Pust, Pathways, pg. 76
\bibitem{GoldmanPust77} Goldman & Pust, Pathways, pg. 77
\bibitem{GoldmanPust73} Goldman & Pust, Pathways, pg. 73
\bibitem{GoldmanPust76} Goldman & Pust, Pathways, pg. 76
\bibitem{GoldmanPust83} Goldman & Pust, Pathways, pg. 83
\end{thebibliography}

My definition of intuition shall, as I stated earlier, be motivated independently of any prior commitment to a particular conceptual scheme. The conceptual scheme to which I will commit shall be motivated by the metaphysical and epistemic concerns of Chapters 2 and 3. Briefly, however, on behalf of G&P, the reason intuition must be about mental states and not the world itself is that the candidates for concept-contenthood that are not mental do not readily gain admission into the intentional realm: Universals, like Platonic entities lack a causal relationship with mental life - how could one be a reliable indicator of the other? Modal equivalences suffer from much the same problem - intuitions are taken as mental occurrences about objects in modal space, and modal space does not lend itself to a relationship with mental occurrences so as the latter can be reliable indicators of the former. Natural kinds suffer from a kind of chicken and egg problem - if intuitions are to judge examples of a natural kind, how is the natural kind itself supposed to be first established? This last question shall be addressed in detail in Chapter 4. This is a
run into the problem that intuition becomes defined in such a way so as to include virtually all thought, and has content strictly about other mental states (they do find, however, that intuition is indeed a basic evidential source - of one’s own thoughts; that, I take it, parallels G&S’s characterization of intuition as psychological data). Their problem is as follows (and a similar problem may bedevil every strictly internalist account of cognition):

A spontaneous, classificational judgment, of the form ‘e is an instance of F’, whereby their take on the content of ‘e’ is a “psychological structure of state that underpins [the] deployment of F as a natural-language predicate” (see above) results in the intuition being of the form “e satisfies my concept that I express through the predicate F [emphasis mine]'60. An intuition is then, really just a spontaneous expression of evidence for a thought - the content of the thought being expressed in the intuition. G&P seem to admit this when they say “Thus, it is not only possible, but almost a matter of definition, that if the concept possessor were fully informed about the relevant features of e, then if e satisfied the concept he expresses through ‘F’, his intuitive response to the question of whether e satisfies this concept would be affirmative; [similarly for the negative] ... in other words a concept tends to be manifested by intuitions that reflect or express its content’61. An intuition is evidence for a thought, the content of which is expressed by that intuition.

How may a thought be different from an intuition? G&P grasp at the idea that intuitions may not be “perfect indicators of their [the concept’s] truth”62. Imperfect indicators would be unfavourable circumstances - such as a case whereby an intuition arises from a mistaken theory about the concept of F rather than from the concept of F itself. Although there certainly may be conceptual wiggle room between one’s theory of a concept of something, and the concept itself as it appears in spontaneous, conscious mental life, that wiggle room certainly doesn’t capture my considered intuitions regarding intuitions. Being less glib, I would point out that mental activity arising from whatever a theory of my concepts may comprise, and mental activity arising from the concepts themselves (whatever that may comprise), do not readily map onto a distinction that would hold between thought and intuition. Further, as I shall argue in Chapters 2 and 3, whatever it is that underpins my judgments, or conceptual classifications - particularly those most amenable to intuitive analysis over conscious deliberation – it is decidedly not an explicit theory.

However, it’s true that an emphasis on the ‘spontaneous’ part of conscious mental life does capture

---

60 Goldman & Pust, Pathways, pg. 83
61 Goldman & Pust, Pathways, pg. 84
something important about intuitions, and it is certainly something that I will retain. It’s also true that the distinction between a person’s theory of a concept and the concept itself may indeed provide for a structured way of distinguishing spontaneous, classificational judgments from any conscious thinking that may be related to that judgment. However, I have three general concerns with adopting this approach. First, unless I adopt and make explicit use of their theory of concepts, I cannot distinguish thought and intuition - the definition is much too broad. Second, I have committed not to have a theory of concepts motivate a theory of intuition; rather, I would like the psychological or cognitive status of intuition to drive an analysis of Standard Philosophical Methodology (which, in turn, drives an analysis of the kind of conceptual scheme SPM is committed to). Third, given the metaphysical and epistemic considerations of Chapters 2 and 3, I am certainly not prepared to adopt a ‘mentalist’ or internalist conception of mind.

That said, there is much of great use here. I take from G&P the (RI) requirement; that is, intuitions must be reliable indicators of the truth of their contents. Changing the terminology somewhat, I shall simply call this a reliability requirement. I also take on board their description of standard philosophical methodology (SPM), and the role in which intuition is meant to play in that methodology. I take the ‘spontaneous mental judgments’ as a key descriptive component of intuition.

The causal relation, which for G&P is not necessary, but the absence of which indicates a reason to doubt that such a mental state fulfils the reliability requirement, I shall elevate to the status of necessity. If there is no causal relationship between the state of affairs which is the content of the intuition, and the mental act of intuiting, then the intuition is not a basic evidential source. I raise the status of causation in this way, simply to state clearly a kind of allegiance to a form of physicalism (which I will explicitly define in Chapter 2), to clearly differentiate this notion of intuition from the all-too-everyday use of the word in which a person may ‘intuit’ that in one’s past life there were goblins dancing on their rooftop. Admittedly, this move shall limit the scope of application of intuition, depending on one’s ontological story. If numbers are Platonic entities, and they have no causal relationship with my cognitive structure, then I cannot intuit anything about them. I do not, however, hold to such an ontological view, but it is not altogether important for the purposes of analysis at hand. I limit the purview of intuition to be about those states of affairs with which one can causally interact.

62Goldman & Pust, Pathways, pg 84
1.4.2 Pust

One of the most important aspects of Pust’s analysis is the broad range of applicability of intuition, from particular cases to much more general intuitions about the suitability of some general principle or of the coherence of some principle with a background theory. I emphasize first the aspects of his approach that apply to that broad range, and then turn to his definition.

1.4.2.1 Particularist versus Generalist Intuitions

Pust starts his analysis by making an important distinction between ‘particularist intuitions’ - those intuitions that are about particular cases, such as the Gettier cases currently under investigation - and ‘generalist intuitions’ - those that are of a more general nature. For Pust, “standard philosophical analysis is a method driven by the evidential appeal to particular case intuitions”, but argues that methods of analysis that attempt to circumvent such appeals are no less reliant on intuition because “those philosophers ... instead appeal to more general intuitions and so also use intuitions as evidence”. The kinds of analyses that “avoid exclusive appeal to particular case intuitions” to which Pust refer are all categorized as one form or another of reflective equilibrium; the best-known and most readily identified example of which comes from Rawls, and his attempts to develop moral theory, but other methodologies can also be shown to fit this profile.

Pust takes note that there are a number of instances in which intuitions appear to be of a more general nature than those of some single, particular case. Many moral theorists, for example, may claim that “a suitably formulated consequentialist moral theory is simple, ‘self-evident’ or intuitive in itself.” Such intuitions may require them to “deny the veracity of many of their conflicting particular case intuitions.” Closer to home, epistemologists may claim that “the closure principle for knowledge (if S knows that p and S knows that if p then q, then S knows q) ought to be accepted because it is quite intuitive. Here, they are, it seems, appealing to an intuition more general in character than intuitions about whether p is known in some hypothetical case.” This seems right. Rawls himself certainly agrees that intuition comes in varying degrees of generality; “People

63 Pust, 2000, pg. 1
64 Pust, 2000, pg. 1
65 Pust, 2000, pg. 1
66 Pust, 2000, pg. 11
67 Pust, 2000, pg. 11
68 Pust, 2000, pg. 11
have considered judgments at all levels of generality, from those about particular situations and institutions up through broad standards and first principles to formal and abstract conditions on moral conceptions.\textsuperscript{69}

The ‘more general’ appeals to intuition are made explicit below in the treatment of reflective equilibrium, but for now let us note that two principles are brought to light. First, “we ought to admit the existence of intuitions at varying levels of generality”\textsuperscript{70} and relatedly, we want to leave open the option that we do not “require treating all intuitions on an epistemic par”\textsuperscript{71}. Pust explicitly endorses a claim we have already seen endorsed by Gopnik & Schwitzgebel and by Wisniewski, namely that “none of what I’ve said should imply that intuitions are treated as infallible or incorrigible evidence. They are not.”\textsuperscript{72}. I fully endorse the claim that there are intuitions at differing degrees of generality, or differing degrees of abstractness\textsuperscript{73} and certainly, the claim that intuitions are fallible.

1.4.2.2 Narrow and Wide Reflective Equilibrium, General and Global Intuitionism

Pust brings together his particularist and generalist notions of intuition into what he calls global intuitionism by way of his analysis of wide reflective equilibrium. For Pust, a possible objection to his view that philosophy is rife with references to intuition would be the claim that the “method of reflective equilibrium (RE) does not require any foundational evidential appeal to intuition”\textsuperscript{74}. Reflective equilibrium may not be as well-defined as the widespread use of the term may suggest, but there is certainly a general pattern to the way in which such strategies are employed. I shall suggest an outline here, following the spirit or main thrust of Pust’s analysis. The method is most often associated with moral theory, following Rawls or some variant of Rawls. Here, I want to characterize reflective equilibrium in the most general way, and emphasize what Daniels\textsuperscript{75} identifies as \textit{wide} reflective equilibrium. I turn now to Rawls and Goodman to illustrate the distinctions between particularist, generalist and global intuitionism and the relation to narrow and wide

\textsuperscript{69} Rawls, 1974, pg. 8
\textsuperscript{70} Pust, 2000, pg. 12
\textsuperscript{71} Pust, 2000, pg. 12
\textsuperscript{72} Pust, 2000, pg. 12
\textsuperscript{73} What one means by ‘generality’ or, in my preferred terminology ‘abstractness’ needs to be made precise. This work is made explicit in Chapters 2 and 3. One of the ways in which this gradient of abstractness can be defined is in terms of the degree of conceptual vs. perceptual content.
\textsuperscript{74} Pust, 2000, pg. 13
\textsuperscript{75} Daniels, 1979, 1980 a and 1980 b.
reflective equilibrium.

In Rawls’ moral theory, one uses a series of individual cases to generate a set of intuitive judgments, and those judgments are then used to derive general principles, the use of which would correspond with, and efficiently summarize, the initial set of judgments. Thus, there is an ordered set of pairs, consisting of a) a set of considered judgments, and b) a set of principles that systematizes the considered judgments of a). Here, particularist intuitions are used to generate principles that explicate those judgments. One can then employ more generalist intuitions about the principles themselves, as to their self-evident suitability or acceptability. At this stage, one is free to assign a “special epistemic status to the particular judgments or to the principles”\textsuperscript{76}. Stopping at this degree of analysis, “we have only achieved narrow reflective equilibrium.”\textsuperscript{77} At this point, one has employed either particularist intuition (if the principles are to cede priority to the individual judgments) or a kind of global intuitionism (if the principles can be adjusted because of a self-evident judgment about their suitability).

Unless one takes into account other theoretical considerations, “It isn’t at all clear why we should want to revise a set of principles which perfectly fit our particular case intuitions unless we find some (less than perfectly matching) principles even more intuitively compelling than our particular case judgments ... it clearly involves no more than admitting more general intuitions into the evidential base.”\textsuperscript{78} This is exactly what Pust calls “… a move from particularist to global intuitionism.”\textsuperscript{79} To emphasize, unless one has independent reasons for modifying the derived principles, one is either implicitly or explicitly intuiting some self-evidential fact about those principles.

An example will help. Goodman discusses how one goes about justifying particular deductive and inductive inferences, namely by referring to the principles of deductive and inductive inference. Those principles must themselves be justified;

“... deductive inferences are justified by conformity to valid general rules, and... general rules are justified by their conformity to valid inferences... this circle is a virtuous one ... rules and particular inferences alike are justified by being brought into agreement with each other. A rule is amended if it yields an inference we are

\textsuperscript{76} Pust, 2000, pg. 23
\textsuperscript{77} Pust, 2000, pg. 24
\textsuperscript{78} Pust, 2000, pg. 22
unwilling to accept; an inference is rejected if it violates a rule we are unwilling to amend.”

In each case, the acceptance or rejection of a rule or of a particular inference, is independently judged. The emphasis can still be placed on particularist intuitions, “if the reason we are unwilling to amend the rule ... is that we recognize such a revision would not allow us to maintain that many other particular inferences we take to be justified really are justified [second emphasis mine].”

On the other hand, it may be glaringly obvious that a principle needs to be amended, contravening many of the individual judgments. Here, I take it that in such a case a strong generalist intuition is relied upon for reasons that are self-evident, or else further reasons are given that have independent support.

In the case that there are independent reasons to adjust the principles, we have moved to an analysis that is characterized as wide reflective equilibrium. Here, I take it that it is not obvious that intuition is still the basis for modification of the principles. It would appear that the level of analysis is sufficiently abstract to warrant the claim that intuition, particularist and generalist, have been left behind. Pust, however, argues otherwise.

The independent reasons (to adjust the principles) must come with some kind of support. They may be supported by some separate (and privileged) set of background knowledge, or by the consideration that the principles should better cohere with some other theoretical or meta-theoretical commitments (which is really just another way of saying background knowledge). The background knowledge itself is either independently derived, or it is itself self-evident. If it is self-evident, then we are back to generalist intuitions. If the background knowledge has independent support, then, for Pust, we have simply pushed the problem back one level, as that independent support will - ultimately – still come down to other particularist or generalist intuitions; “so long as those [background theories] require justification by appeal to particular intuitions, we will not have escaped the evidential appeal to particular intuitions.”

There is thus now an ordered triple; a) a set of intuitive judgments, b) a set of principles that can be derived from those judgments, and c) a set of relevant background theories.

The means by which one makes the ordered triple cohere, in such reflective equilibrium, is the

---

79 Pust, 2000, pg. 22
80 Goodman, 1983, pp. 63-64
81 Pust, 2000, pg. 18
important point I am bringing out here; I do not claim to have *proven* either that, all philosophy comes down to wide reflective equilibrium, nor that, all wide reflective equilibrium is intuitionist in nature. What I claim to have done is motivate reasons to think it may be the case. Pust has characterized a quite broad notion of intuition; as a cognitive skill, it is quite general in scope in that it encompasses both imagined particular cases of semantic assent, and much more general ‘coherentist’ intuitions in which one intuits the proper fit of quite abstract and conceptual items. I will rely on as broad an applicability of intuition as possible, including the more abstract and general conceptual realm, and so find some support here. Nothing important hinges on whether I have logically proven some necessary characteristic of philosophical analysis *in general* - that would require a much, much larger treatment than I am prepared to give.

Pust’s motivation in relating global intuitionism to wide reflective equilibrium is to emphasize the pervasiveness of intuition in philosophy. I take his point, but my motivation in emphasizing the same line of thinking is to be able to develop an analysis of intuition that can draw on the epistemic continuum I outlined at the outset. Perceptive judgments will be the most readily analyzed, since that is what the connectionist literature emphasizes, and I want to lay out a way of motivating the application of perceptive principles to more conceptual judgments. That ‘intuition’ is the same term - and has much the same import and character, from individual cases to more general ‘coherentist’ concerns - will be a part of that motivation.

Pust, like myself, believes that “one species of psychological state undergirds philosophical practice and therefore serves as the ultimate source of evidence for much philosophical theorizing,” and he takes as his project the task of giving an “account of the psychological states the content of which are used as [intuitional] evidence for philosophical theories”

Our projects are quite similar. However, he narrows his view considerably from the outset as to what kind of mental phenomenon that psychological state is meant to include. It does not, for Pust, include “what is meant by the more colloquial use of the term ‘intuition’ or what disciplines other than analytic philosophy mean by the term ... for example ... in empirical psychology or common sense lore.”

It is a common strategy in philosophy to limit the jurisdiction over which one proclaims responsibility, as there are commensurately fewer angles of attack from which one must defend oneself, and Pust has ensured that his analysis remains tightly focused on a single ‘target [which

---

82 Pust, 2000 pg. 27
83 Although Pust does claim to prove the point, it is not necessary for me to establish so strong a claim
84 Pust, 2000, pg. 30
85 Pust, 2000, pg. 30
is] philosophical intuition”

At the risk of developing a position that is less easily defended, I shall part with Pust on this point. I shall claim that a common psychological state undergirds a much broader view of intuition, one that includes both perceptive and ‘purely conceptual’ states (such as counter-factual or ‘modal’ analysis) as well as the more colloquial notion (with some requisite qualifications). My broader notion shall assuredly come at the cost of opening up more angles of attack, but equally at the benefit of gaining a much greater degree of explanatory power for intuition qua ‘psychological’ or cognitive state.

1.4.2.3 The Nature of Intuition as “Intellectual Seeming”

Pust takes as an important methodological principle that he is “trying to determine what ‘having an intuition that p’ involves from a first-person point of view” and ultimately, “will argue that intuitions are a distinct kind of psychological state with their own ‘intellectual’ phenomenology”. In this strategy, I agree with Pust, as I too, shall start from a phenomenological perspective. The starting point for Pust is, in following Bealer, to describe intuition as an “intellectual seeming” (rather than as a “judgment”, as he had done in Goldman & Pust, above). An ‘intellectual seeming’ is a conscious episode in which “when S has the intuition that p, p ‘just seems’ true to S”. There are two important points about this ‘intellectual seeming’ that Pust develops in detail. First, he develops a means of demonstrating how belief and intuition come apart. Second, he analyses the kind of necessity that is involved - that is, he asks: if the intellectual seeming could not be otherwise, or is basic in some way, what is the nature of the necessity of the seeming? I shall disagree with Pust about the nature of the necessity, although I shall use it as a starting point in my own considerations about the nature of the necessity relation involved in intuitive judgments.

Pust develops a final definition that is as follows:

[A1] S has a rational intuition that p IF AND ONLY IF (a) S has a purely intellectual experience, when considering the question of whether p, that p; and (b) at

---

86 Pust, 2000, pg. 30
87 Pust, 2000, pg. 31
88 Pust, 2000, pg. 31
89 Pust: “… my work ... owes a significant debt to Bealer’s writings”, Pust, pg. 32. I shall follow Pust’s analysis directly, as the analysis is indeed quite similar to Bealer (1987, 1993, 1996), but Pust takes it a step further.
90 Pust, 2000, pg. 32
t, if S were to consider whether p is necessarily true, then S would have a purely intellectual experience that necessarily p.

P, here, is propositional in character. In limiting himself to philosophical intuitions, Pust leans toward terminology like ‘rational intuition’ and ‘purely intellectual experience’. As I shall part with him on the scope of the applicability of the term in question, I shall not question nor analyse this part of his definition (however, I take some support from the ‘purely intellectual’ character, as I will certainly want to include such phenomenology in my definition).

I’m not entirely sure, nor does Pust make it explicit, why he has shifted from his spontaneous judgment starting point in Goldman & Pust (shared by a number of other philosophers)\textsuperscript{92}, and shifted to an intellectual seeming, particularly as he is limiting himself to those philosophical intuitions used in philosophical analysis. Such intuitions are invariably judgmental in character, whether particularist about individual cases, or generalist in theoretical coherence or suitability. Earlier, Pust himself talks quite often of intuition as judgment\textsuperscript{93}. I am not arguing that Pust is wrong, but I am unsatisfied with his shift, which seems entirely driven by his allegiance to Bealer\textsuperscript{94} and his (mistaken\textsuperscript{95}) nature of the necessity of intuition, rather than by his self-proclaimed desire to “determine what ‘having an intuition that p’ involves from a first-person point of view” (quoted above). I point this out because I follow Pust in using a first-person phenomenology to drive an initial characterization, and spontaneous judgment is one clear manifestation of that phenomenology.

1.4.2.4 Intuition and Belief

Pust demonstrates that belief and intuition come apart when one digs a little deeper into the phenomenological character of the intellectual seeming. Belief that p is not sufficient for having an intuition that p, nor is belief that p necessary for having an intuition that p. There are a number of

\textsuperscript{91} Pust, 2000, pg. 32 
\textsuperscript{92} See Williams (2004, pg 109); “What are called ‘intuitions’ in philosophy are just applications of our ordinary capacities for judgment.”, Gopnik and Schitzgebel (pg 77); “Because of this important role of intuitive judgments ... intuitive categorization judgments ... manifested through intuitive judgments ...”, etc, etc. 
\textsuperscript{93} “...the bulk of the intuitive judgments ...”, pg 18, “... considered judgments at all levels of generality are evidential ... we have here ... nothing more than global intuitionism ...”, pg 22, “... cast doubt on the veracity of our intuitive judgments regarding particular cases”, pg 25, etc. 
\textsuperscript{94} “I find most of Bealer’s claims about the nature of intuition ... quite plausible”, Pust, footnote, pg. 32 
\textsuperscript{95} See Section 1.4.2.5
ways, and examples, of pointing out this distinction.

To show that belief is not sufficient for intuition, take the following example. A mathematical friend, who has earned high accolades in her profession, proves to you, step-by-step, that a particular mathematical theorem is true. Not being a mathematical genius yourself, the truth of the theorem remains unintuitive to the extreme, yet it is perfectly conceivable that you now believe it. Thus, “Beliefs are not seemings in [this] sense of the term.”96, and belief is not sufficient for intuition. Conversely, some mathematical proposition may seem true, one may intuit the truth of the theorem. However, the trusted mathematical friend now comes along and proves it false. Here, you may still intuit that the theorem is true - that is, it may intellectually seem to you to be true, but you no longer believe it. Thus, belief that p is not necessary for an intuition that p.

For Pust, philosophical paradoxes follow much the same pattern. In such a paradox, “one is confronted with intuitive propositions which lead by intuitively acceptable reasoning to an intuitively unacceptable conclusion.”97 After the paradox has been noted, “one is likely to suspend belief in all of the relevant propositions without their diminishing in intuitive force.”98 Indeed, after going to the effort of resolving the paradox, and dropping belief in one (or more) of the propositions, those propositions “retain[s] a fair amount of intuitive plausibility”99

Here, Pust points out that there appears to be “the intellectual equivalent of a perceptual illusion”100. We are asked to consider the Mueller-Lyer illusion (although there are any number of other such illusions that bring the point home). Our perceptive systems give us an unequivocal ‘seeming’ of the two lines being of different lengths, a seeming that persists even when one has gone and measured the two lines; “I believe quite firmly that the two lines are of identical length. The appearance that p, in such a case is impenetrable101 by belief and even though I have the belief

---

96 Pust, 2000, pg. 32
97 Pust, 2000, pg. 33
98 Pust, 2000, pg. 33
99 Pust, 2000, pg. 32
100 Pust, 2000, pg. 33
101 Although, interestingly, it seems to me that once the lines have been measured, there are ways of ‘squinting’ or ‘looking more closely’ at the illusion to have it begin to drop away. It seems to me that there is demonstrated here (anecdotally, I admit - and I welcome the reader to do the same) some indication that a strongly held concept can be consciously brought to bear on a perceptive experience, changing (somewhat) the quality of that experience. Thus, in some cases at least, the appearance that p is not impenetrable by belief, although it strains effort to do so. There are other ways of consciously changing perceptive experience, such as depth perception and the quality of shapes, that artists have often referred to, and even tried to extrapolate to their canvases, learned presumably by focussing attention in particular ways at particular objects within their perceptive field.
that not p, p still perceptually *seems* true.”\textsuperscript{102}

I certainly agree that intuition and belief are distinct, and I agree that these kinds of examples indicate the conceptual separation between the two. However, a ‘seeming’ is not a ‘judgment’, and as my analysis will hinge on intuition as a judgment, rather than a seeming, in my view the two concepts also come apart along a different axis. Beliefs are not judgments, but more like a set of background theories that *inform* judgments.

1.4.2.5 Intuition and Necessity

There is a kind of necessity in the *intellectual seeming*; “a genuine intuition must be such that the truth of p strikes us as *necessary*”\textsuperscript{103} Pust analyzes this necessity relation (provided originally by Bealer) as:

\begin{align*}
[B1] \text{At t, S [rationally] intuits that p IF AND ONLY IF at t, it intellectually seems to} \\
\text{x that p and also that necessarily p.}\textsuperscript{104}
\end{align*}

I take it that the spirit of this necessity relation is the *foundational epistemic character* of the intellectual seeming. The state of affairs intuited just could not *seem to be* any other way. This seems right, and the spirit of this sort of epistemic necessity I would certainly want to retain; “the sort of state at issue admits, like a pure phenomenal colour, of no further analysis.”\textsuperscript{105} This agrees with the kind of intellectual illusions analogous to the perceptual illusions outlined in the previous section - such *seemings* do indeed seem hold to fast to their conceptual character in some important way.

However, it does not appear that this is the kind of necessity that Pust has captured. The definition ends up being quite austere, as the necessity relation is placed such that it would limit the kinds of intuitions that one could have, if they are to be genuine intuitions, to have conceptual content that deals exclusively with *a priori* truths. ‘Necessarily p’ points to the idea that p is a *logical* necessity. Pust is quite aware of this, and asks “Are all philosophically evidential intuitions phenomenologically like seeing that p *must* be true? Are there not situations in which p seems true

\begin{flushleft}
\textsuperscript{102} Pust, 2000, pg. 33. \\
\textsuperscript{103} Pust, 2000, pg. 35 \\
\textsuperscript{104} Pust, 2000, pg. 36
\end{flushleft}
to us, but not necessarily so? ... [we are concerned with], among others, epistemic, moral, metaphysical and semantic intuitions." Pust attempts to address this issue by

"amend[ing] B1 to not require that the intuited proposition seem necessarily true at t, the time of the intuition, but to require instead that at t, S be disposed (under appropriate conditions) to have it seem necessarily true." 

This analysis still seems to equivocate on the purview of the necessity relation. While I understand Pust to interpret this definition as having the necessity relation apply to the epistemic state of the intuiter - that it simply couldn’t seem any other way to her - I don’t see it as the only possible interpretation of the above. It could still be interpreted as implying that the state of affairs that is referred to by the intellectual experience occurring at t, is necessary. Pust has shifted emphasis to an “occurrent seeming necessary,” and in so doing, is trying to emphasize the epistemically necessary character of the seeming, not the necessary character of the p that is the object of the seeming. Pust makes a final move to remove the relation altogether, so that [A1] above is modified to lose b) section entirely. I think such a move unwarranted, as the intuition about the epistemic necessity of intuitions is a good one.

So I propose a friendly amendment, that the necessity relation be explicitly epistemic in character. Such a treatment would capture the line Pust is pursuing in emphasizing that the seeming is such that it can be likened to a basic phenomenal state like a ‘pure phenomenal colour’ or a perceptual illusion such as a Mueller-Lyer illusion. I shall incorporate this notion of epistemic necessity in my definition.

---

105 Pust, 2000, pg. 36
106 Pust, 2000, pg. 37
107 Pust, 2000, pg. 39
108 Pust, 2000, pg. 50
109 Note that by ‘epistemic necessity’ I mean something quite different from the term as used by Kripke in Naming and Necessity: for my purposes this term means something like I-could-not-have-judged-otherwise. This point shall be developed presently.
1.5 Intuition, A Definition

In motivating my definition of intuition, I note two points. First, I would like the definition to be driven by phenomenological and psychological analysis that is (as) independent of other philosophical commitments (as possible), particularly in terms of any prior commitment to a theory of concepts. I follow Pust in allowing “armchair psychology”\textsuperscript{110} into my analysis. I start with asking the questions ‘What is the phenomenology of intuition – what it like to experience intuition?’ and ‘What makes intuition distinct from other psychological phenomena?’ I do this to ensure that the cognitive analysis of intuition that follows on from my definition is not tainted by a prior commitment to a theoretical position I would then like to turn around and criticise. In other words, if I am to use a cognitive analysis of intuition to refute logical atomism, for example, I had better not have prior commitments to some competing theory of concepts buried in my definition.

Second, at the risk of opening myself to more angles of attack, I would like my definition to be as broad as possible, capturing both the austere philosophical ‘intellectual seemings’ used in philosophical analysis, as well as the more common uses of the term in psychological literature and quotidian use. Here, I certainly part ways with Pust. My purpose is to make the most of what I will claim is a common underlying cognitive mechanism. A methodological consideration I made plain at the outset was the intention to derive principles that appear quite well documented at one end of the epistemic continuum (ranging from perception at one end) and apply them in a way that is more than analogical to the other (to intuitive modal analysis at the other). Thus, a broad range of epistemic activity will fall under my definition of intuition.

That said, neither my broad definition, nor the validity of the epistemic continuum, nor the methodological consideration just mentioned, stand without independent support. My definition captures, not just an everyday commonality of reference, but something that does indeed - I claim - have a common underlying cognitive mechanism (at least as described from the relevant functional perspective). This claim, and the epistemic continuity supposition, find independent support throughout Chapter 3, as does the methodological consideration (both of which, of course, are related to the claim that there is a common underlying mechanism).

I start with what appear to be two quite different everyday notions of intuition, and I shall develop a

\textsuperscript{110}Pust, 2000, pg. 31
definition of intuition that encompasses the two. One is the *hunch* associated with an everyday usage of the word such as “My intuition tells me he is dishonest” - call this sort of intuition **H-Intuition**. The second is the sort of *obvious*, self-evident truth that fills the role of ‘basic evidential source’ in SPM. Call this sort of intuition **O-Intuition**. The distinction between the two shall hinge on the degree of consciously introspective evidence available to support the intuition. There are, as one might expect, intermediate cases - but there shall nevertheless be a clearly defined distinction between cases of each type. My strategy is to develop an analytic definition that reconciles these two notions under one mental capacity.

### 1.5.1 H-Intuition; Reliable Hunches

I start with a notion of intuition that is normally associated with a ‘hunch’, a kind of awareness of some aspect of the world. This awareness may often require quite complex (if tacit) sets of inferences. The defining feature of this kind of mental event is that any ‘figuring out’ that is required is unconscious. There is no conscious, careful or computation-like thinking involved\(^1\). The hunch appears to the intuiter as a bare and unmediated awareness of some aspect of the world. The awareness is immediate, and is of a more rapid character than the thoughtful analysis of rational, conscious knowing. The first working definition, then, I take from the psychological literature on the subject;

**[H1] H-Intuition:** “direct, immediate apprehensions without conscious rational analysis”\(^2\)

I take two points from [H1]. ‘Apprehensions’ are a kind of judgment. To apprehend some state of affairs is to apprehend it *as being* some state of affairs. To apprehend that a person is angry, is to see that person *as angry*. To see that person as angry is to make a judgment the person is angry, to *categorize* their mood as one of anger. So, to apprehend is to judge, and to judge is to judge *that* some state of affairs obtains. So here, I first substitute Goldman & Pust’s notion of ‘spontaneous mental judgments’ for ‘apprehensions’ and, secondly, I (again following Goldman & Pust) qualify

---

\(^1\) The important distinction between *contingent* and *in principle* unavailability of a ‘thinking’ or ‘computational’ process (and the nature of the relationship between inference, thinking and computation) I leave aside for now, since I am concerned primarily with a phenomenologically-driven definition. *Contingent* unavailability would be a case in which computation/inference took place, but it is simply not made available to consciousness. *In principle* unavailability would be a case in which there no computation/inference took place that *could* be made available. This issue will be given some precision in section 1.5.3, and shall dealt with in detail in Chapters 2 through 4.

\(^2\) Myers, Pg. 75
those judgments as being comprised of ‘singular classificational propositions’, although I shall use slightly different terminology.

Note that these two qualifications narrow the original scope considerably. Definition [H1] really only eliminates conscious, rational analysis from the spectrum of mental phenomenon. Memory, for example, would classify as a direct, immediate apprehension, as would the apprehension of a kind of ‘inner monologue’. In the first case, I directly apprehend a memory and in the second, I apprehend the sound of my own inner voice. Neither of these phenomena should qualify as intuition, and the ‘spontaneous mental judgment’ serves to eliminate these phenomena. I take note of the criticism that apprehending an inner voice is a ‘spontaneous mental judgment’ of a kind, one in which it is judged that some particular word or phrase occurred. The relationship between language apprehension and intuition is complex, and will be developed in greater detail, but for the present analysis, language apprehension is not a ‘hunch’; it is more like O-Intuition, an example of intuiting the obvious.

Further, by emphasizing the ‘mental’ character, I also mean to eliminate the normal, unexamined state of perception. When one perceives an object, one perceives the object *qua* some object. When I perceive an apple, I perceive that object *as an apple*. There is intentional content in the perception, namely, that the object is an apple. The intentional content is rarely made explicit, in describing (to oneself or others) what one sees, but is implicit in the unconscious organizing of our everyday experience into one of *objects*. I would like to eliminate this kind of perception from H-Intuition, as it – like the case of language apprehension – is not the kind of thing one has a hunch about - it is more akin to intuiting the obvious.

Also note that specifying that the judgment is to be of propositional form eliminates a number of physical skills, such as that of the professional quarterback, whose acquired skill in throwing the ball includes a vast number of calculations and judgments involving wind-speed, distance, etc., which are quite clearly not consciously deliberated. While such skilful activity will turn out to have a close relationship - from a cognitive perspective - with the intuition we are concerned with, it does not capture either the phenomenological or psychological character I am after here.

Refining [H1] then,

[H2] **H-Intuition**: direct, immediate and spontaneous mental judgments ‘that p’, in propositional form, that occur without conscious rational analysis.
With a linguistic basis for the content of the judgment, the H-Intuition has enough symbolic and clearly defined conceptual content to be placed (in some appropriate way) within an epistemic framework, within what Sellars calls the ‘space of reasons’. The Sellarsian ‘space of reasons’ rests on the necessity of being able to provide additional linguistic support – reasons – when possessing an intentional item of any kind. Entertaining some mental item as an item – the entertainment being intentional in character – necessarily entails that other linguistic propositions sit in some complex web of epistemic support to that intentionally-described item. H-Intuitions ‘that-p’ then, following Sellars (1997), require that there the intuition exist within a linguistically formulated framework that provides support for that proposition.

I take the Sellarsian point – however, I want to distinguish a psychological or phenomenological perspective from the philosophical perspective – being in the philosophical space of language does not necessarily put you in the phenomenological space of reasons. What I mean by this qualification of the typical Sellarsian picture shall be made clear in Section 1.5.3, but briefly - I am able to provide reasons for my hunch-like intuitive judgments to a greater or lesser degree, sometimes not at all, even though those judgments are of a propositional character and to have made those intentional judgements I must already be in the space of language. The content of the hunch is propositional in form – and thus sits in the Sellarsian space of reasons – but may or may not sit in the phenomenological space of reasons. This distinction is meant to emphasize the ‘direct’ and ‘immediate’ qualifiers. Retaining the ‘direct’ and ‘immediate’ qualifiers really just re-affirms the phenomenology of the judgment as appearing fully fleshed out in intentional terms upon arising within one’s conscious purview, and arising (seemingly) unrelated to a set of supporting reasons. However, I do not deny the more basic Sellarsian point that one must possess linguistic skills to have intentional items in the first place.

At this point, a further refinement is brought to bear, namely, that of Goldman & Pust’s Reliability Indicator, which I remind the reader is the requirement that, in order for a mental state to count as basic evidential source, the state in question must be “[a] reliable indicator[s] of the truth of their contents”113. There are two sorts of H-Intuitions that could be readily identified a this point, and the Reliability Indicator is meant to eliminate one. The first is the sort of reliable, expert judgment as exemplified by the chess-master. Having spent a considerable time using rational analysis to identify positions of weakness and strength on the chess-board, and spending many formative chess-years consciously pondering what may be the best move on a given board, there comes a
point at which such conscious deliberation disappears and the chess-master can not only ‘see’
complex contexts such as “weak king side” but rarely has to consider more than a handful of
moves to come up with the best move within those contexts. I will have much more to say about this
kind of expert judgment presently in my analysis of expert systems, but for the time being let us
note that experts such chess-masters, experienced medical diagnosticians and car mechanics, have
intuitions that are more than mere ‘hunches’, they are indeed reliable indicators of the truth of their
contents.

The reader need not be fully convinced, at this point, of the reliable expertise demonstrated in the
intuitions of such experts, but should be convinced enough to note the distinction between such
intuitions and those unreliable, tenuous intuitive judgments (almost always, it seems, given in
retrospect!) of the sort indicated by statements like “I intuited my long-lost aunt would call
tonight!” or “I intuit that your dead father wants to talk to you”. I want to eliminate from my
definition this particular colloquial use. Unless it can be empirically demonstrated that the
intuitions in question fulfil some form of the Reliability Indicator requirement, then such mental
phenomena are to be eliminated from discussion. So, our definition now becomes:

[H3] H-Intuition: direct, immediate, spontaneous and reliable mental judgments
‘that p’, in propositional form, that occur without conscious, rational analysis.

The best examples of this sort of intuition, then - as distinct phenomena from intuiting the obvious -
are the developed intuitions of the expert. Hunches, to be reliable, are learned over time (I contend).
A fisherman, after a life at sea, may reliably intuit the weather. A doctor may intuit a diagnosis, a
psychologist a deeply-hidden motivation, and a chess-master a small set of potentially best moves.
Learning over time is not the only way to develop reliable hunches, however, nor are hunches
learned over time guaranteed to be reliable. I could, in theory, take a pill that renders my brain such
that my hunches are reliable. I could also, over time, discount positive qualities of alternatives I
decided not to take, in a kind of retrospective rationalization of my choices. That said, I take it that
learning over time, gaining expertise with experience, is a quite typical method of generating the
requisite reliability.

I have remained neutral thus far in terms of the question as to whether there is introspectively
availability evidence for the intuition, and have emphasized that neutrality in drawing a distinction
between the philosophical and phenomenological characterizations of the ‘space of reasons’.

113 Goldman & Pust, 2002, pg. 74
Before I turn to that question, I now outline what I mean by O-Intuition.

1.5.2 O-Intuition; Intuiting the Obvious

The obvious, self-evident or often common-sensical nature of O-Intuitions is what I take to be their defining feature, and I take it that it is the property that motivates the high epistemic status bestowed upon them by many philosophers. This self-evident nature is what would enable such intuitions to be regarded as a ‘basic evidential source’. Keeping this feature in mind, along with some of the refinements made in the previous section, we have the following definition:

\[
\text{[O] O-Intuition: direct, immediate, spontaneous judgments ‘that p’ (in propositional form) that appear self-evident and obvious.}
\]

O-Intuition encompasses a lot of what one would normally call common sense. How do I know that space is three-dimensional? How do I know I exist? How do I know that two objects can’t share the same spatio-temporal location? How do I know my telephone is not conscious? Each of these questions have common-sensical answers, and I am claiming that much of this everyday, obvious knowledge can be regarded as comprising intuitive judgments. O-Intuitions are often not made explicit, and sit as a kind of background knowledge that informs a lot of our thinking/theories, and certainly permeates our day-to-day activities as we cope with getting on in the world. It shall turn out that this sort of knowledge, the knowledge which we pretty much take for granted and which is here defined as O-Intuition, is actually the most difficult to codify and is a defining feature of human (vs. machine) intelligence.

I am explicitly allowing the kind of judgments characteristic of everyday perception. Seeing an apple \textit{as an apple} is such a judgment - it appears self-evident that the object is an apple. Although one may be able to produce all sorts of reasons \textit{why} the object is an apple, that it is an apple nevertheless appears as an obvious, simple fact. At bottom, I claim, is the feeling that it’s an apple \textit{just because it seems that way}. Any supporting reasons are retrospective in nature, and do not match the temporal phenomenology of the O-Intuition itself. Also, however much supporting evidence one may be able to give, it will ultimately feel unsatisfactory. The phenomenology of everyday perception is such that the judgments appear foundational. It will turn out that this ‘foundational’ reading of O-Intuition has a lot of common ground with Fodor’s informational atomism – see Chapter 4.
That said, I do not want to imply that perception and intuition cannot, if pressed, be torn apart. By including ordinary perception within the purview of intuition I am emphasizing my commitment to the epistemic continuum; it being a continuum, however, it is surely subject to more carefully conceptualized divisions and distinctions. If I were to allow perception to be cleaved from intuition, it would be so with respect to the issue of justification. A key component of intuition – more so than for perception, since the very definition of intuition rests upon it – is the idea that my reasons for judgement run out. Justification always, on my view, runs out at some point, and the question – and the distinction that I would allow between perception and intuition – is: how soon do my reasons run out? For intuition (by definition) they run out immediately. For perception, I may well be able to continue to provide, for some time, reasons for my judgement.

To this issue - of justification, reasons and evidential support - I turn now in detail. I am trying to get at a characteristic of intuition that is best described as something like: ‘I know this fact, but I don’t really know why I know it.’ It seems to me a key aspect of intuition that it comes without a concomitant explanation of why one knows what one knows.

1.5.3 Introspective Support and Epistemic Necessity

I address now the important issue of whether or not there appear to introspection supporting reasons that have led to the content of the intuition; that is, whether or not one is able to give reasons as to why the judgment takes the form that it does, and relatedly, whether those reasons actually play any inferential role in the formation of the judgment. For brevity, let me call the introspectively available reasons and/or the actual underlying inferential process ‘cognitive support’. I mean to tie the question of cognitive support to the notion of epistemic necessity outlined previously in the Pust analysis.

Epistemic necessity is a statement about the epistemic feel of the judgment - whether it seems one could have judged otherwise. Epistemic necessity, as I have defined it, is a property of the phenomenological experience of the intuition, it is only indirectly related the content of the intuition. A priori statements are logically necessary and counterfactual statements are often assumed to be have truth conditions supplied by necessities of various sorts (nomological necessity to conform with physical law, for example); the necessity relation involved in these cases is dependent on the content of the statements themselves. Here, epistemic necessity is defined such that the necessity
relation involved is dependent on the phenomenological character of the intuition, not its content –
the relevant question is whether or not it may seem to the intuiter that the content of the intuition
could have been otherwise.

Since I am here concerned with the phenomenological character of intuition, not the actual cognitive
process underlying that phenomenology (as I shall be in Chapter 5), I remain completely neutral at
this stage as whether there really is some unconscious, inferential and intentional cognitive process
that takes place and can be brought into conscious inspection or whether there only appears to be
such a process. I here separate two distinct sort of questions. The first is whether or not there
really is an unconscious process that could be brought to the light of consciousness –
corresponding to a contingent/in-principle unavailability of underlying cognitive support – and the
second is whether or not there appears to be that underlying cognitive support. There are thus four
distinct possibilities:

    a) There is an underlying inferential process, but no conscious access to it.
    b) There is no underlying inferential process, and no conscious appearance of one.
    c) There is an underlying inferential process, and conscious access to it.
    d) There is no underlying inferential process, but there appears to be access to one.

In cases a), c) and d) there is cognitive support, and in case b) there is none.

The question of the existence of such a mechanism will be deferred until I am explicitly tackling the
issue of the underlying cognitive mechanism itself. At the phenomenological level, the
implementation is irrelevant; what is relevant, at this level, is the character of the mental experience
associated with the intuition.

The point here is to simply ask the following question of the intuiter: can you give adequate and/or
complete reasoning behind your intuitive judgment? Can you provide us with all of your reasons?
Call the set of reasons available to be given the Supportive Set.

By ‘adequate’ and ‘complete’ I make an important point - call this set of reasons the Fully
Supportive Set. The Fully Supportive Set is defined to be sufficient to account for that judgment
(although it is not defined to be necessary, as there may be many distinct conjunct Fully Supportive
Sets). By sufficient here, I mean only phenomenologically sufficient - that is, the intuiter feels as if
they have given a complete explanation of the judgment. It does not matter whether or not there is a
logical inconsistency, or epistemic gap, or lapse of reasoning in the Fully Supportive Set - what matters is that, from the intuiter’s perspective, it feels as if a complete set of reasons is given. The expert’s intuition - a chess master, for example - may well be able to, in retrospect, provide a Fully Supportive Set of reasons as to why the best move was to move his queen three squares to the left. That may or not have been the actual unconscious reasoning, and it may or may not be an entirely satisfactory explanation from an epistemic perspective, but if it feels like a complete explanation to that chess-master, then it is a Fully Supportive Set.

To give rough-and-ready reasons why, reasons that may indicate a tendency toward that judgment and not another, is to give a Partially Supportive Set. Some explanatory support is given, but it does not feel sufficient - although the intuition remains as strong as ever. The intuiter has the phenomenology of there being some missing gap in the explanation, regardless of the strength of the intuition. There would be a tendency to give the Partially Supportive Set and then just say something like “But that’s it, it just seems to be the case that p!” Philosophical intuitions often fall into this category. Given a counter-factual scenario, an intuitive judgment is elicited, and it is the nature of such constructed examples that a Fully Supportive Set cannot be given - that’s the point of referring to intuitions. A Gettier case may invoke a judgment that such-and-such is not a case of knowledge, and perhaps a Partially Supportive Set can be given, but it is almost always the case that the explanation stops and the argument rests on the content of the intuition itself.

This analysis admits of degrees. Since the distinction between Partially- and Fully- Supportive Sets rests on the phenomenological sensation of explanatory satisfaction, then the distinction between the two is not logically discrete. Depending on the phenomenology associated with the Supportive Set, depending on the degree to which the Supporting Set is a satisfying explanation, then to that degree it is a Fully Supporting Set.

It is true that there is an equivocation here between an after-the-fact rationalization, that may or may not reflect the actual decision-making process, and the reasoning that is actually responsible for the judgment. But just as I am not here concerned with actual cognitive processes versus apparent cognitive processes, I am equally unconcerned with actual reasons versus reasons constructed in retrospect. Here, all I want to do is get at what degree of epistemic support is introspectively available to support the judgment. Questions relating to actual cognitive activity shall wait until Chapter 3 – Recognition.

What I am trying to do here is make distinct two sorts of H-Intuitions. In one sort, one can provide
a Fully Supportive Set of reasons as to the content of the judgment. Call these cases Quasi-H-Intuition. In another sort, only a Partially Supportive Set is available. Call these cases Real-H-Intuition. This analysis admits of degrees, as noted above. Pedantically: To the degree of satisfaction provided by the explanation contained in the Supportive Set, then to that degree the Supporting Set is a Fully Supporting Set, and correspondingly, to that degree, the intuition is to be classed as a Quasi-H-Intuition. To the degree that the explanation contained in the Supporting Set is unsatisfactory, then to that degree the Supporting Set is a Partial Supporting Set, and correspondingly, to that degree, the intuition is to be classed as a Real-H-Intuition.

I now want to relate epistemic necessity to Real- and Quasi-H-Intuitions. Here, I am trying to make explicit the kind of necessity that Pust wanted to emphasize in his ‘intellectual seeming’, the necessity that indicated that the content of the judgment is unalterable, even in the face of counter-evidence. Just as one cannot ‘unsee’ the Mueller-Lyer illusion, having measured the lines, one cannot ‘un-intuit’ the judgment. Epistemic necessity, then, is also a phenomenological property, and is a property of the feel of the intuition. It too, comes in degrees, and I want to say it is a measure of a kind of immediacy of the intuition.

Until now, this section has been largely definitional, with an emphasis on developing a conceptual framework by which to analyse the phenomenology of an intuition and its relation to an introspectively available justificatory framework. What I now want to argue is:

To the extent that an intuition is a Real-H-Intuition, that is, to the extent that only a Partially Supportive Set of reasons is available, then to that extent the intuition will appear as Epistemically Necessary. Relatedly, to the extent that an intuition is a Quasi-H-Intuition, that is, to the extent that a Fully Supportive Set of reasons are available, then to that extent the intuition will appear as Epistemically Contingent.

The argument goes like this. In cases in which one has a complete set of justifications, one would have an accompanying feeling that the content of the judgment could have been otherwise. One could easily imagine that such-and-such may not have been the case, or if reason number three had not prevailed, then the state-of-affairs would be differently judged. With explanation comes other possibilities. The intuition may appear as an immediate, direct, judgment but is a judgment that appears epistemically contingent. That is, one can see how it could be otherwise - the intuition is not foundational, in an important way. A chess-master intuits the best move, but because reasoning that would be fully supportive of that move is available as introspective evidence, the intuition is felt
to be merely an epistemic short-cut, made up of isolable steps that form a rational chain of events - any one of which could have been otherwise. Just as the state of the game is contingent on past moves, the placement of the players, the ability of the opponent, so too, the intuitive best move is contingent upon a series of cognitive steps. Quasi-H-Intuitions, I claim, do not come with a feeling of necessity since with the degree of supporting explanation comes a degree of feeling that the judgment is merely a short-cut, formed of more foundational epistemic steps.

On the other hand, if one cannot provide sufficient reasons, if one cannot fully explain the judgment, the intuition has, I argue, a different epistemic feel. In such cases - of Real-H-Intuition - the intuiter has the feeling that the judgment could not have been otherwise. The judgment, as an unexplained explainer, sits as a kind of foundational epistemic episode - at least from the phenomenological perspective of the intuiter. The series of explanatory questions - why? - comes to an end, and with the end of inquiry comes necessity. In philosophical analysis, the point of developing counter-factual examples is to develop just such a situation. In a Gettier case, the instance is judged not to be knowledge, and perhaps some reasons can be given as to why not, but at it’s core, such methodology is designed to elicit the response “It just seems that way to me”. If the intuition elicited was a Quasi-H-Intuition, then the appeal to intuition could be eliminated in favour of the discrete, isolable epistemic steps that comprise the Fully Supportive Set.

Note that I have explicitly related epistemic necessity to epistemic foundationalism - indeed, I do hold such a view. To the extent that an intuition is epistemically necessary, it is foundational – it is a judgement for which no other reasons can be provided. This follows from the fact that an epistemically necessary intuition cannot be explained (or at least fully explained) and so must remain as a part of the explanation itself. In that sense, it is foundational.

While I believe I am correct in the analysis of the accompanying ‘feels’ - epistemically necessary or contingent - not a lot rides on the claim going through in its entirety. The claim of epistemic necessity could, in a pinch, be regarded as definitional in character: I am defining Quasi-H-Intuitions as ones for which one can give reasons, and I am calling such intuitions epistemically contingent. I am defining Real-H-Intuitions as ones for which one cannot give sufficient explanation, they appear ‘bare’ in a way that Quasi-H-Intuitions do not, and I am calling such intuitions epistemically necessary.

So, I define Real-H-Intuitions as:
[H4]: **Real-H-Intuitions**: direct, immediate, spontaneous and reliable mental judgments ‘that p’, in propositional form, that appear epistemically necessary; that is, they appear as if they could not have been otherwise.

I note for completeness that a Partial or Full Supporting Set of reasons may themselves be instances of Real- or Quasi-Intuitions.

### 1.5.4 Intuition writ Large: H- and O-Intuitions as Pattern Recognition

What I will now argue is that Real-H-Intuitions are really the same thing - from a phenomenological perspective - as O-Intuitions, and provide some examples of what I mean. These examples shall be worked up in great detail in Chapters 2 and 3.

That Real-H-Intuitions and O-Intuitions amount to the same thing, from a phenomenological perspective, should already be clear. Recall **O-Intuition**: direct, immediate, spontaneous judgments ‘that p’ (in propositional form) that appear self-evident and obvious. The epistemic necessity associated with Real-H-Intuition amounts to the same thing: the intuited judgment could not have been otherwise; it is thus, from an evidential perspective, self-evident - for where else could evidence come from? I admit that Real-H-Intuitions may not be obvious in the restricted sense that it may take work, training or experience with the phenomena in question to be able to see that phenomena in the light of a Real-H-Intuition. ‘Weak king side’ may not appear self-evident to a novice. However, the point is that *once the relevant experience is gained*, then the Real-H-Intuition does become obvious. Similarly, I don’t suppose to an infant it is obvious that some object is a compass, but it does become obvious over time. Henceforth, I shall adopt the definition of intuition as that of Real-H-Intuition given above in [H4], as the self-evident nature of intuitions is given greater precision in the analysis of epistemic necessity.

I have allowed that O-Intuition include the perception of objects. I have also indicated that some justification could be given for such classificatory acts: the apple is an apple because it is round-ish, green, has a stem, and - yes - tastes like an apple. However, I also take it that such reasons run out, in everyday experience - “It’s an apple because it - *seems to be / looks like/ appears to me as - an apple.*” I take it perception of objects has this fundamental character. Perception of objects is a kind of intuition, and is epistemically necessary. I can’t help but see a cup as a cup, a tree as a tree or a happy face as a happy face. *My sympathetic reading of Fodor’s informational atomism is*
driven by my commitment to epistemic necessity. See Section 4.4.4.

As the epistemic continuum asserts, perception and cognition are distinct, but overlapping, activities. The least conceptual perceptual items may be primary colours such as red, less conceptual an apple or tree, up to the conceptually loaded like instances of ‘graceful’ or ‘knowledge’. A chess-master seeing ‘weak king side’ is perceiving a conceptually loaded state, indeed one that takes a good deal of training to see. Sports enthusiasts perceive different states of affairs when watching their favourite game than do amateurs - to see a good centre trap in hockey is to see a complicated state of affairs that requires a good deal of concepts and conditioning. Gettier cases are presumably fully conceptual, as they are imagined counter-factual instances in modal space, upon which one must pass judgment as to the applicability of the term ‘knowledge’.

Each of these cognitive acts I take as being instances of intuition. I shall argue that each of these acts is a kind of pattern recognition. Judging a situation to be of a particular kind, such as an instance of ‘red’, or ‘weak king side’ or ‘knowledge’ is really recognizing the situation as fitting in to a kind of pattern.

What I want to develop, then, over the course of the next four chapters is the following idea:

**Intuition is a kind of pattern recognition.**

---

114 I am sympathetic to, but remain neutral on, the Sellarsian notion that even such epistemically primary items, such as colours, are not perceived as colours until one enters into the space of reasons. The point here is that, whatever conceptual framework one needs in place in order to perceive, such items are simply the least conceptually loaded.
Chapter Two

2. PATTERN METAPHYSICS

We have seen in the previous chapter that of particular concern is the analysis of a concept; how is it to be analyzed - what is the proper role of the folks’ intuition in that analysis? A concept has been likened to a pattern, and the intuitive judgments regarding the applicability of a concept to a particular case is a case of pattern-recognition. The validity of this move – equating patterns to concepts – shall be brought out throughout this chapter and I ask the reader to withhold judgment on that move for the present.

I turn now to the question of what, exactly, a pattern is: what is it that is being recognized, and what supports the equating of concept with pattern? My objective is to build a framework by which one can speak in an orderly way about the sorts of patterns that compose those objects to which we make reference in our natural language, a framework that has a number of degrees (or axes) of complexity by which to speak of increasingly interesting patterns. I shall show that attempts to split the notion of pattern from recognition are unsuccessful, aside from some artificially simple examples. Thus, the metaphysics of pattern that is developed remains deeply enmeshed with the epistemology of pattern, and the definition of ‘pattern’ retains an operational flavour.

I shall focus on the work of three philosophers; Dennett, Haugeland, and Dreyfus (I return in detail to Dreyfus in Chapter 3).
2.1 Metaphysics of Patterns – The Landscape

The framework I shall use to discuss the complexity of patterns is defined in terms of three axes along which the degree of complexity may be viewed as varying. The complexity of a pattern is to be measured (in a very loose, non-quantitative way) by way of its assignment to one of three distinct Levels (I, II and III) of increasing complexity, that assignment being governed by the pattern’s: i) amenability to algorithmic description, ii) context-dependency, iii) mind-dependency. These axes are not logically independent, and there are a number of ways in which they may be related. There are five sorts of patterns that I shall use as indicators of points along these axes; mathematical objects (like geometrical shapes, or numbers), natural kinds (maple trees and tulips), artifacts (chairs and churches), gaming kinds (Life and chess) and of course the most interesting sort of pattern, intentional kinds (knowledge, hopes and beliefs).

Amenability to algorithmic description (AAD) – the question of whether the identification conditions associated with patterns of some determinate type can be specified in the language of traditional computing methods - is of central importance here, since a central aim of this chapter is to build a view of the metaphysics of pattern that places severe limits on what sorts of computation can lead to a pattern’s identification: symbolic, number-crunching machines will only be capable of identifying those lower-level patterns of Level I and possibly Level II. Context-dependency is to be regarded as supplementary to the question of AAD, in the sense that it provides additional reason to think that some patterns are not AAD; context-dependency then, is meant to provide grist for the anti-AAD mill. Mind-dependency enters the picture in that some patterns – those that are most interesting and relevant to the larger discussion at hand (intentional terms, for example) – are complex enough to be thought of as requiring the presence of an interpreting mind to identify them.

I shall differentiate the operational notion of pattern from the notion that patterns are discernable-in-principle: patterns may well be discernable-in-principle – that is, they may exist in the world, ready and available for discernment independent of whether or not minds are present – even though they require a mind to pick them out. Hence, the AAD analysis places constraints on what sorts of cognitive equipment may discern certain patterns of interest, and hence eliminates the possibility of the identification of those patterns in some mind-independent manner, but it is not meant to necessitate a descent (an ascent?) into a kind of idealism, in which those patterns are not really there but are merely projections or expressions of the interests of the perceiver. Patterns may thus be discernable-in-principle, but without a non-Turing sort of mind to apprehend them, could not be
picked out.

The sorts of complexity that I develop leads to an irreducibly operational definition of pattern: a pattern will be what a pattern-recognizing creatures takes to be a pattern. With some relevant caveats I shall avoid vacuity or idealism; those caveats being that the creature is to be regarded as embedded in a socio-linguistic community and the acts of recognition are to be put to the test of enabling that creature’s pragmatic success in negotiating its way around in the world. That to which our concepts refer - which correspond to these patterns - are not, on this account, the sorts of things that are amenable to being described by way of independently defined necessary and sufficient conditions. That said, intuition – the ability to recognize a pattern of interest – may act to provide an entry point to generating a set of such conditions; those conditions, then, turn out to be dependent on the interests and abilities of a cognizing agent.

2.1.1 A Quick Note on Physicalism

Underlying this view is a commitment to a kind of physicalism, a view that I shall not explicitly defend, but present as a kind of premise. I take it that were there no matter – and by matter I mean whatever basic constituents are the building blocks of atoms, be they quarks, gluons or charm(ing) electrons – there would be no patterns to speak of at all. I take it as an empirical question of basic physics as to whether those basic constituents are themselves patterns, or whether they constitute objects. Call those basic building blocks – whatever they are - Ontology I. Ontology I has a kind of primacy – hence my commitment to physicalism – in that all other patterns, all other objects to be defined in one’s ontology, would not exist were it not for the prior existence of Ontology I. “No matter, never mind” strikes me as an apt motto for this brand of physicalism. Ontology II may be thought of consisting of whatever other ontological furniture one wants to posit; chairs, tables, brain states, intentional items, or even numbers and triangles. Items in Ontology II owe their existence to Ontology I - regardless of their complexity, epistemic status, durability or temporal relations - simply because I take it that all of the furniture of Ontology II would disappear if that of Ontology I were to do so. Items furnishing Ontology II are patterns of Ontology I.

This is a very weak brand of physicalism, at least as presented in the form of this premise. To be

\[115\] I understand that abstract items, such as numbers and triangles, are some of the most controversial items when it comes to even this mild sort of physicalism. However, while acknowledging that there are complex issues surrounding mathematical objects in particular, I shall press on with this assumption. There is a position on this issue which I am comfortable endorsing, while not explicitly defending, which is – funnily enough – intuitionism as best espoused by L.E.J. Brouwer; see Brouwer in Benacerraf, P. & Putnam, H., 1964.
clear, I am making no claims (at this point) about the primacy of Ontology I entailing, precluding or denying any sort of reducibility or identity either in practice or principle, be it a law-like type-type reducibility or even a much weaker token-token identity. No prior assumption is being made about how the relationship holds between Ontology I and Ontology II, other than Ontology II consists, at bottom, of some arbitrarily complex swirling soup of Ontology I. My position will gain clarity (by way of commitments to various forms or possibilities of reduction) throughout this chapter, but for now I assume only a most basic commitment to physicalism. Indeed, one of the main points I bring out in the hierarchy of complexity shall be that most patterns of even mild interest are not reducible either in practice or in principle; but that relationship is derived, not assumed!

Patterns then, however simple or complex, are to be regarded as patterns of quarks (or gluons, or whatever) and that’s all.
2.2 Dennett and Real Patterns

Dennett’s primary concern is to defend (or better – articulate) the sort of ontology associated with the intentional states that become apparent when adopting his ‘intentional stance’\(^{116}\); the intentional stance is the way we approach other persons in order to best make sense of their behaviour. The ‘stance’ is taken in order to maximize the predictive power one has of the system(s) under observation, and – certainly for the systems that are people - the best way to predict and explain behaviour is to adopt the stance that those systems do indeed have intentions (and beliefs and desires and so on). The intentions themselves (as shall become clear presently) are to be regarded as observable, empirical patterns of behaviour. Thus, if the best predictive power regarding a system is gained by adopting a stance toward it \textit{as if it had intentions}, then that system is to be regarded as having those intentions\(^{117}\). The patterns of behaviour which become apparent, and of which we make sense, in adopting the intentional stance toward our fellow people – patterns which correspond to essentially all of the elements of folk psychology – exhibit, in Dennett’s view, a sort of “mild realism”\(^{118}\). This position is contrasted with a number of others, three of which I shall articulate; Fodor\(^{119}\) who is an “industrial strength Realist [and for whom] beliefs … would not be real unless the pattern dimly discernable from the perspective of folk psychology could also be discerned … as a pattern of structures in the brain”\(^{120}\), Churchland’s “eliminative materialism, which denies the reality of beliefs altogether”\(^{121}\) and “Davidson’s regular strength realism”\(^{122}\). It is in setting out the case for his mild realism that Dennett articulates - what is for me - the main lesson; that these sorts of patterns not only have a special kind of mild realism associated with them, but the realism is of a mild sort \textit{because of the kind of complexity that they exhibit}.

\(^{116}\) I shall assume the reader is somewhat familiar with the main force of Dennett’s work, particularly that of The Intentional Stance, (Dennett, 1998). Other stances, such as a design stance, are also possible. A chess-playing computer is best understood and explained if one adopts the stance toward that computer that is has been designed to play chess.

\(^{117}\) One of the most prevalent criticisms of this view is the charge that it is \textit{by virtue of other observers adopting a stance} toward a system (or creatures) that a system is to be regarded as having intentions, rather than by some intrinsic fact about that creature’s mental life. The intentional stance then, faces criticism that it is overly instrumental in character. This criticism, however, does not account for the distinction that I shall bring out throughout this chapter: a pattern may be discernable-in-principle – and thus \textit{there to be seen} – regardless of whether or not it necessitates an observing creature, adopting a stance, to see it.

\(^{118}\) Dennett, 1991, pg. 30

\(^{119}\) See Chapter 4 for a full exegesis of Fodor.

\(^{120}\) Dennett, 1991, pg. 42

\(^{121}\) Dennett, 1991, pg. 30

\(^{122}\) Dennett, 1991, pg. 30
The intentional patterns by which we predict and explain each others’ behaviour are “intermediate regularities [that exist] between the regularities of planets and other objects ‘obeying’ the laws of physics and the regularities of rule-following … systems.”¹²³ These patterns are complex enough that, for Dennett, it is possible in principle for two observers to apply (perceive) two distinct patterns (sets of intentional states) for the same set of interpreted events (set of behaviours); “The choice of a pattern would indeed be up to the observer … [yet there is] … abundant evidence that our allegiance to folk psychology as a predictive tool can be defended in coldly objective terms.”¹²⁴ Dennett takes Quine’s¹²⁵ indeterminacy thesis – the claim that there will always remain an indeterminacy as to the meaning of a person’s linguistic expression, given the totality of their behaviour - to be radical indeed, yet he still grounds all such patterns in empirical, observable data and puts those patterns to the test of predictive power. For Dennett, it is the ‘noisiness’ of the patterns that enables this indeterminacy.

The predictive power of folk psychology (and its associated intentional terms) is what is of paramount concern, for Dennett, since that is what enables the empirical grounding of the associated intentional items, even though we use folk psychology for all sorts of other reasons; “Folk psychology helps us to understand and empathize with others, organize our memories, interpret our emotions, and flavor our vision in a thousand ways, but at the heart of all these is the enormous predictive leverage of folk psychology. Without its predictive power, we could have no interpersonal projects or relations at all; human activity would be so much Brownian motion.”¹²⁶ To predict is to fill in a pattern, to extend that pattern or to fit it as an element in a larger pattern. Seeing, or being cognizant of those patterns, is the basis of the intentional stance, and it is only by adopting such a stance toward others that we are able to perceive these patterns.

Adopting a particular stance, as we shall see however, is dependent on a creature’s ability to take that stance, which in turn, is dependent on that creatures physiological makeup (endowed by evolution), as well as their own development (endowed by their own efforts and cultural-historical position). The multiplicity of possible interpretations derives in large part, then, from differences in the epistemic state of the observer – here I bring out the idea that mind-dependency is introduced as one measure of a pattern’s complexity. What Dennett insufficiently analyzes, however, is the distinction between the idea that there are a multiplicity of patterns that are there to be identified - they are each discernable-in-principle, regardless of the epistemic state of the observer - and the idea that an observer picks out a particular pattern - which is nevertheless indeterminate.

¹²³ Dennett, 1991, pg. 43
¹²⁴ Dennett, 1991, pp. 49-50
¹²⁵ See Quine, 1960, Ch. 2
2.2.1 Barcodes and Noise

The first simple visual pattern Dennett provides is a series of nine squares that alternate between black and white, dubbed ‘bar code’. Each square is nine pixels by nine pixels, and so another way to think about it is nine lines of nine sets of alternating black and white pixels. There are varying degrees of noise inserted into bar code, from very little (in which the larger squares are perfectly clear) to 50% noise (in which the larger squares are indistinguishable from the noise, since at this point the noise has served to eliminate the distinction between background and foreground). The important point that gets brought out here relates directly to the presence of the noise.

I take it that the real world is messy, that objects from cows to maple trees, from people to tulips are each distinct – and it is our ability to categorize distinct, individual objects and place distinct, unique events under concepts that is of interest here. Without noise of some kind – at least as variance from some prototypical example of an object - objects falling under the same concept would be each identical to each other, just as each bar code would be identical, and just as any intentional item would be identical. We would live in a world of (something akin to) Platonic forms. This is manifestly untrue. A distinction must be noted, however, between that which is random and that which is mere variance. Individuals may vary, but may not vary randomly. I shall put aside that distinction for now, however, as Dennett’s point can be brought out without it. Note that Dennett does not mean for noise here to be regarded as mere interference, artificially inserted to make a trivial point about dots and squares, but as motivating the discussion as being about more than just bar codes (it’s about the messy real world). I here limit my discussion to the example itself, and extend the principles from the simple bar code to complex intentional items in Section 2.2.3.

In defining what it is to be a pattern, particularly in the presence of noise, Dennett follows a well-worn path in making use of information theory, but in so doing he slides from an operational definition of pattern (that it is a candidate for pattern recognition) to a more formal notion (that it is a mathematically-defined signal that can be compressed). A pattern – that which is to be extracted as the relevant similarity between each more or less noisy bar code – is “in the root case… a

---

126 Dennett, 1991, pg. 29
127 I put off for the present the question as to what, precisely, ‘noise’ is supposed to be – see Section 2.2.4 for details. I am not entirely satisfied with Dennett’s easy (and, for him, key) moves between ‘noise’ and ‘signal’, but do think that his main points still stand.
candidate for pattern recognition and more formally, this “discernibility-in-principle” is made precise by comparing it to that which is random; “A series ... is random if and only if the information required to describe (transmit) the series accurately is incompressible.” So a pattern is something that is discernable in principle, which distinguishes it from random signals, and that notion is made precise with the use of information theory and its quantifiable notion of compressibility; “A pattern exists in some data – is real – if there is a description of the data that is more efficient than the bit map, whether or not anyone can concoct it.” I shall leave discussion of the vacillation between the operational and formal notions until the end of this section.

White noise (made up entirely of random elements) has no pattern because there is no way to compress it, a perfect bar code has a clear pattern in that it can be re-described using much less information (‘nine squares side-by-side, each nine pixels in length, alternating black and white’, suitably encoded, for example). Bar codes with intermediate amounts of noise are intermediate cases – if the noise is to be captured as well. Seeing through the noise to what we are calling ‘bar-code’ (the perfect, noiseless version), is a kind of pattern-correction; here, the noise-less bar code would serve as the paradigmatic case, the Platonic-like form of the pattern.

There are two important points Dennett brings out from this example. First, he points out that one person’s noise may well be another person’s signal; there is no way in principle to decide which interpretation of the data is the correct one – although this argument, as we shall see, can be mitigated by repeated exposure to the signal. However, if the noise is – as Dennett’s example implies – random, then it cannot, by definition, be itself a signal. I address this idea in more detail in the next section. Second, and relatedly, the pattern that is distinguished may have any number of underlying explanatory processes (or methods by which those patterns are caused).

---

128 Dennett, 1991, pg. 32
129 Dennett, 1991, pg. 32
130 Dennett, 1991, pg. 32
131 Dennett, 1991, pg. 34
132 A defining mark of digital systems – such as a Turing machine or a genetically reproduced code – is that they are designed (or evolved) to eliminate noise, to reproduce patterns without noise, or to be able to deal in strictly these ‘Platonic-like’ entities; “Digitality is actually one of Plato’s discoveries. In essence it is the substitution of a three-term relation for a two term relation in the real definition of resemblance. ... if we compare each thing to a paradigm and not the the latest in a series of copies ... reproductive errors will not accumulate.” (DeSousa, 1991, pg. 6).
Remembering that predictability is the crux of the matter, it is entirely possible that one person (or creature) may detect a particular rhythm in the noise itself (or more importantly, what is being interpreted as noise by the other) and use a more complicated – but in some cases more accurate – calculation to predict what the signal might look like in the future. Prediction is a kind of filling-in of the relevant pattern that is recognized. One person may see one sort of pattern with one kind of noise (say just the ‘pure’ bar code with a degree of noise in the form of what appear to be random bits thrown in), and another a different pattern with commensurately different noise (say a complicated interpretation of a rhythm that appeared as noise to the first person). If each is able to make accurate predictions, and thus have their interpretation borne out in “coldly objective terms”\(^{133}\), then each is able to claim that they have detected a real pattern. What is key here, is that each person is capable of showing that their interpretation holds up under repeated application. A part of my definition of intuition, given in the first chapter, is that the judgments must be ‘reliable’, that is, bear up under the scrutiny of repeated exposure to a test of accuracy; in this case, the ability to make accurate predictions must bear out under some number of iterations, and thus be well grounded in inductive evidence. If one person’s interpretation does not, and holds only for a limited sample, it must be said that one interpretation is of the signal, another of noise (that temporarily showed a discernable pattern).

A question remains: what is noise here? Dennett wants noise to be something that can potentially be interpreted as a signal. This is key to his point about one person’s noise being another’s signal. For it to be a signal to someone – no matter what power of prediction they possess – it cannot be random. A signal is – by definition – something that is not random, even by Dennett’s own informational treatment of signal given above. Yet the noise Dennett introduces into ‘bar-code’ is, for all intents and purposes, being treated here as if it were a random element. If it is indeed random, then Dennett’s point about the real indeterminacy – the real choice - of pattern does not stand up to the scrutiny of real predictive power – it must merely appear as random to one person, but not the other. This is why a signal cannot be random; it is something the perception of which allows for predictive power. I submit a friendly amendment to his bar-code example; the noise is not random, but can appear to be random to some, and not to others.

Noise simpliciter, then, I take to be random, as it is in Dennett’s example. That cannot be a pattern as it is an incompressible signal. What we need is a different notion of noise that can be relativized

---

\(^{133}\) Dennett, 1991, pg. 50
qua noise to an observer - since it is a property of the observer, for Dennett, that renders a part of a signal as either noise or pattern. I turn now to this question: what is it about an observer that makes it, relative to them, whether or not a signal is noise or pattern? What is it then, that differentiates the sorts of patterns that emerge for a given observer?

The choice as to which pattern one perceived in interpreting the data is dependent on a number of things, but primarily it is a matter of the inter-related issues of expediency, interests and abilities; “they may perceive different patterns in [the data], but since we can have so many varied interests and perspectives, these differences do not all count as disagreements. Or in any event they should not.” 134 The bar-code example is meant to encapsulate an operating principle by which we see patterns, and identify those which fall under concepts, and the choice between interpretations (assuming those differing interpretations stand the test of reliability over time) is meant to reflect characteristics of the viewer. The difference between an epistemological distinction – those patterns we are able to see – and a metaphysical distinction – those patterns which are available to be seen – remains in place. The point being developed here is to establish the multiplicity of patterns available from the metaphysical perspective.

In terms of expediency, the trade-off is between ease of computation and accuracy of prediction. For example, we employ concepts everyday to make shortcuts in pragmatically dealing with the world around us - we do not compute the trajectories of minute particles. Undoubtedly, computing all the trajectories of what I have called Ontology I would be the most accurate way to predict the precise nature of some future state, but at the expense of a computational complexity of such a degree as to be absurd. Our concepts are expedient short-cuts that compress the swirling mass of the bit map of Ontology II into useable, interesting chunks with which we are capable of dealing.

Our interests and abilities are related to expediency, and this is not just a matter of the colloquial use of the terms which refer to those things to which we choose to apply our attention and with a view toward which we develop our abilities (like playing the guitar), but also the deeper sense of the term which refers to the sorts of things that creatures of our type (or our culture) have developed a special kind of deep relationship. Bees have such a relationship to flowers, and the (invisible to us135) ultra-violet patterns that cover them; “Other creatures with different sense organs, or different interests, might readily perceive patterns that were imperceptible to us. The patterns would

134 Dennett, 1991, pg. 35
135 At least until we developed technology to extend our perceptual abilities
be there all along, but just invisible to us.136 We have such a relationship with language – words on the page – and apples: it is expedient for us to be able to see apples, and interpret language, and we have interests and abilities that enable that expediency; ‘These ‘design decisions’ are typically not left to us to make by individual and deliberate choices; they are incorporated into the design of our sense organs by genetic evolution. The product of this design evolution process is what Wilfrid Sellars calls our manifest image.”137 Thus, Ontology II is brought into focus out of Ontology I by our collective conceptual and perceptual apparatus; our visual and nervous systems, our folk physics, folk psychology, everyday objects, and by what I want to call intuition.

In terms of the more colloquial meaning of ‘ability’ though, as something gained through effort (like playing the guitar), our knowledge base is also often brought to bear on what patterns we are able to perceive; “Differences in knowledge yield striking differences in the capacity to pick up patterns. … Expert chess players, unlike novices, not only know how to play chess; they know how to read chess – how to see the patterns at a glance.”138 Developing expertise opens up new possibilities of pattern.

Note then, that there are two kinds of mind-dependency involved here. First, many patterns we pick out are what I shall call globally mind-dependent: the kind of mind (and perceptive system) we have as an evolved species picks out which of the (infinitely) many ways of interpreting Ontology I is suited to our interests, and so which patterns are made readily available for us to perceive. Such patterns are mind-dependent in the sense that they are useful organizing principles by which we navigate our way around the world. Apples are food to us, and the moon a part of a romantic setting. Second, the patterns available are also singly mind-dependent: the kinds of patterns made available to any one individual are dependent on the degree of knowledge, etc. that they bring to bear on the situation. As an extreme example, imagine the pattern of ‘rising inflation’; this surely, is dependent on the sorts of minds we have, and is some particularly human (perhaps the highly-trained sub-species of homo economicus) interpretation of the vast signal of our bustling surroundings – talk about signal compression!

So the patterns we perceive are not a unique interpretation of the swirling mass of data around us, and the particular interpretations we actually make are highly dependent on what we as cognizers bring to the scene. I will have much more to say about recognition in the next chapter, but note that

136 Dennett, 1991, pg. 34
137 Dennett, 1991, pg. 36
138 Dennett, 1991, pg. 34
– at least for Dennett – the visual example of bar code brings out some deep principles about the multiplicity of patterns available in a given data set. The multiplicity of patterns are all real. This ‘deep indeterminacy’ shall be discussed further in Section 2.2.4: The Indeterminacy of Real Patterns, and the implications are there extended to directly confront the more complex patterns associated with intentional terms.

2.2.1.2 The Indeterminacy of Underlying Explanatory Processes

One could respond – quite reasonably, it seems to me – to all this talk of multiplicity of patterns by postulating that there is some unique, underlying causal process responsible for producing that pattern and hence, there is a preferred pattern about which one can say it is the most real, or the one that is really there. The other patterns are interesting interpretations, but as far as ontology goes, one preferred pattern can be picked out by providing an explanatory mechanism that accounts for its production; “But how could the order be there, so visible amidst the noise, if it were not the direct outline of a concrete orderly process in the background?”

Fodor, of course, makes this exact point in arguing for a language of thought; the best explanation of intentional patterns of behaviour is that there are concrete mental particulars that cause that behaviour. I shall talk extensively of Fodor and the language of thought in Chapter 5, and my extended response to his question is to be found there, but I note here that it is to this type of argument (postulating an underlying causal mechanism) that Dennett provides a response in the form of an extension of the bar code example.

Dennett extends his bar code example by providing an alternate means by which the large-square, visual pattern of bar code may be generated. The large black and white squares which in the previous section were “created by a hard-edged process (ten black, ten white, ten black …) obscured by noise, while the [alternate version] were created by a process almost the reverse of that: … a pattern created by a normal distribution of black dots around means at x = 10,20,50,70, …”

The exact alternate process need not be detailed, suffice to say it is a completely different algorithm and the patterns look roughly the same. Dennett; “My point is that even if the evidence is substantial that the discernible pattern is produced by one process rather than another, it can be

---

Dennett, 1991, pg. 43
Dennett, 1991, pg. 44
rational to ignore those differences and use the simplest pattern description (e.g. bar code) as one’s way of organizing the data.”

Dennett emphasizes that one cannot dodge the lessons of the bar code example by “peer[ing] behind the scenes at the program [he] devised to create the frames” as there could be any number of underlying causal processes responsible for the visual pattern. This is not, to me, an entirely satisfactory response, and I shall outline my misgivings in Section 2.2.5 Dennett: Cautions & Conclusions.

2.2.2 Life, Chess and a Multiplicity of Perspectives

A second example Dennett uses to set up some basic principles of patterns is the cellular automata game of Life. I shall assume the reader is familiar with this game. One of the most intriguing facts about these cellular automata – aside from the almost alarming array of large-scale visual patterns one sees dancing around on the screen – is “the proof that a working model of a universal Turing machine can in principle be constructed in the Life plane!” Since a Turing machine can compute any algorithm, such a machine could be set up to run a chess program. Such a computing Life board would be large; “the whole construction, a self-reproducing machine incorporating a universal Turing machine, would be on the order of $10^{13}$ pixels.” The sense of scale is important here, as the point that Dennett is going to bring out is that patterns visible from one perspective are simply invisible from another; “Perspective would shrink the pixels of a self-reproducing pattern to invisibility. … A self-reproducing pattern would be a hazy glow, like a galaxy.”

---

141 Dennett, 1991, pg. 44
142 Dennett, 1991, pg. 43
143 For those who are not, very briefly: Life is a computerized game, of sorts, in which a grid of cells are given initial values, corresponding to ‘alive’ or ‘dead’. The values evolve over time, in discrete steps, given some simple algorithm in which only adjacent cells may affect the values of other cells. For example, a cell with less than 4 live neighbours may survive, more than 4 and it dies, and exactly 4 it becomes alive (it’s born). As the system evolves, tremendously complex patterns emerge (the status of alive or dead is associated with a colour), with various entities flying around the screen, appearing and disappearing and so on. The large-scale phenomena that one can see, are of course, wholly dependent on local causation at a scale that is often much smaller than the one at which activity – or patterns – emerge. Given the unambiguously local nature of causation in Life, it can be quite startling to see the preponderance of large-scale phenomena emerge, move and change over time. ‘Giders’ are patterns that sustain their integrity as they move across the screen, an ‘Eater’ is a pattern that destroys other patterns that run into it (such as a Glider) while maintaining their integrity, and so on. Life is something that really must be seen to be appreciated!
144 Dennett, 1991, pg. 40
145 Dennett, 1991, pg. 40
146 Poundstone, quoted in Dennett, 1991, pg. 41
relevant patterns is a matter of being “fixed on an interpretation scheme”\textsuperscript{147}, or adopting the correct ‘stance’.

Recall that in shifting from one perspective to another in the bar code example, in which case one was (theoretically) able to interpret the noise as a part of the signal, there was a tradeoff between computational complexity and expediency. Being capable of detecting complicated rhythms in the noise, and taking that information as a part of the signal in order to make predictions about the pixel values at later points in the signal, may lead to the capacity to give more accurate predictions about individual pixels, but that accuracy comes at a cost of great computational complexity. Patterns are, for Dennett, epistemic short-cuts that allow us to make efficient and pragmatic predictions, predictions that are simply not available to us at other scales or degrees of detail. Adopting a stance is picking a perspective, or fixing an “interpretation scheme”, that allows for large-scale activity to become apparent.

In the Life/chess example, looking at individual pixels “would almost certainly be unilluminating”\textsuperscript{148}, but once one adopts the stance that the Life board is a Turing machine, “enormously efficient ways of predicting the future of that configuration are made available”\textsuperscript{149}. A first step is to go from pixels to gliders and other basic patterns, a next step is to shift “from an ontology of gliders and eaters to an ontology of symbols and machine states”\textsuperscript{150}, at which point one becomes capable of “predict[ing] its future as a Turing machine”\textsuperscript{151}. Going one step further, to an even larger scale of events one “can shift to an ontology of chess-board positions, possible chess moves, and the grounds for evaluating them; then, adopting the intentional stance toward the configuration, one can predict its future as a chess player”\textsuperscript{152}, one that is performing intentional actions.

Dennett is emphasizing that some patterns are only visible from a particular perspective – not only from a perspective that can encompass enormous numbers of pixels over some similarly enormous number of generations, but also from the perspective one gains by figuring out that the machine is to be interpreted from a particular point of view – that it is playing chess; “real but (potentially) noisy patterns abound … there for the picking up if only we are lucky or clever enough to hit on the

\textsuperscript{147} Dennett, 1991, pg. 41
\textsuperscript{148} Dennett, 1991, pg. 41
\textsuperscript{149} Dennett, 1991, pg. 41
\textsuperscript{150} Dennett, 1991, pg. 41
\textsuperscript{151} Dennett, 1991, pg. 41
\textsuperscript{152} Dennett, 1991, pg. 41
right perspective. They are not visual patterns but … intellectual patterns”\textsuperscript{153}. In adopting the chess stance, one gains an enormous leverage in being able to predict the configurations of vast chunks of the Life board – chess concepts are here the epistemic shortcuts that trade off computational complexity for pragmatic ability. One is compressing the signal by interpreting patterns in the signal.

The lessons are not limited to chess and Life, of course, since this example meant to provide insight into our own swirling universe of Ontology I, the signal of which is massively compressed; the “scale of compression […] in the Life example […] is stupendous. But the scale of the savings is really no greater in the Life world than in our own.”\textsuperscript{154} When I throw a brick at someone, or bat my eyes at a pretty girl, or wave ‘hello’ to an old friend at the airport I do not perform calculations at the level of Ontology I, nor at the level of physiology, and so on. I use the patterns of intentional psychology.

2.2.3 Lessons: The Indeterminacy of Real Patterns

Each of the two previous examples was used to generate principles about patterns in general, although Dennett is concerned with patterns that emerge from adopting the intentional stance. It is intentional patterns that I shall focus on here, retaining the lessons learned from the visual patterns given above.

The first point to note is, of course, that the patterns one perceives are necessarily noisy. Patterns of behaviour are what Dennett calls intermediate regularities, and – rather than being the outcome of stable and concrete internal order such as a language of thought – they “could be … the statistical effect of very many concrete minutiae producing , as if by a hidden hand, an approximation of the ‘ideal’ order”\textsuperscript{155}. Such intermediate regularities – like the patterns in a game of Life that are only visible from a certain perspective and only at a certain scale – exist “between the regularities of planets and other objects ‘obeying’ the laws of physics and the regularities of rule-following (that is, rule-consulting) systems.”\textsuperscript{156} The source of such regularities – these ‘intermediate regularities’ - is handed off to evolutionary pressures, and as such, they are more than mere regularities - they

\textsuperscript{153} Dennett, 1991, pg. 41
\textsuperscript{154} Dennett, 1991, pg. 42
\textsuperscript{155} Dennett, 1991, pg. 43
\textsuperscript{156} Dennett, 1991, pg. 43
are what “one would expect self-designing systems to ‘discover’ in the course of settling into their patterns of activity”\(^\text{157}\).

Interpreting others, and applying intentional concepts to explain and predict their behaviour requires an “ineliminable use of idealization”\(^\text{158}\), and because of the associated ‘noise’ – which as we have seen above, may be noise to one but signal to another – leads to the real possibility of there being rival interpretations of the same data. That is, there may be two distinct interpretations of the same behaviour. Dennett draws a strong conclusion from this fact – he adopts a strong version of Quine’s Principle of Indeterminacy of Translation, in which not only the “choice of pattern is up to the observer”\(^\text{159}\) but “no deeper fact of the matter could establish that one was a description of the individual’s real beliefs and the other not.”\(^\text{160}\). This, of course, does not happen much of the time but such “radical indeterminacy is a genuine and stable possibility”\(^\text{161}\).

Dennett is taking a stronger stand than Davidson concerning the indeterminacy of intentional states, for whom these are “competing descriptions of the same reality”\(^\text{162}\) – hence Dennett’s moniker of “Davidson’s regular strength realism”\(^\text{163}\). For Davidson, the indeterminacy of intentional states is much like different temperature schemes - they really are just different interpretations of the same underlying reality of thermal intensity. For Dennett, there simply is no fact of the matter, there is no primacy of some particular underlying reality. Davidson’s take on indeterminacy is, then, for Dennett, “not the shocker it is often taken to be; in fact it is well-nigh trivial”\(^\text{164}\). All Davidson is saying is there are different ways of seeing things, but there is a determinate way things are, and what Dennett is saying is that is no determinate way that things are (as far as intentional states go).

The problem of indeterminacy cannot be solved by reverting to a lower scale, or a more detailed examination of the brain, because in lowering one’s perspective to that level one forgoes the intentional stance (and the patterns that stance enables you to perceive) and takes on instead some sort of ‘brain stance’, at the level of which none of the behaviour-predicting patterns would have been observable in the first place (just as in the Life game, one cannot perceive the chess patterns at the level of gliders or individual cells). The apparent precision we have in communicating,

---

157 Dennett, 1991, pg. 43
158 Dennett, 1991, pg. 48
159 Dennett, 1991, pg. 49
160 Dennett, 1991, pg. 49
161 Dennett, 1991, pg. 48
162 Dennett, 1991, pg. 46
163 Dennett, 1991, pg. 30
164 Dennett, 1991, pg. 46
predicting and understanding each other derives from language and that precision can be illusory, and not reflective of an underlying unity or universality; “The process that produces the data of folk psychology … is one in which the multi-dimensional complexities of the underlying processes are projected through linguistic behaviour, which creates an appearance of definiteness and precision, thanks to the discreteness of words.”¹⁶⁵.

This is all anathema to someone like Fodor, of course. For Fodor, the best explanation of the order we observe in behaviour - and, indeed, the source of our ability to communicate tout court - is derived from, and subservient to, the precision of the syntax of brain states. Order at the personal level could not be anything other than a “direct outline of a concrete orderly process in the background”¹⁶⁶. Dennett; “Fodor and others have claimed that an interior language of thought is the best explanation of the hard edges visible in ‘propositional attitude psychology’”¹⁶⁷. For Fodor, the beliefs and other intentional items that are the source of the patterns seen at the level of behaviour and speech “would not be real unless the pattern dimly discernable from the perspective of folk psychology [that noisy intermediate regularity] could also be discerned (more clearly, with less noise) as a pattern of structures in the brain.”¹⁶⁸ Dennett’s argument is meant to capture the idea that there may be any number of underlying sources of the observed regularities, and also any number of competing descriptions or interpretations of those regularities.

2.2.4 Dennett; Cautions & Conclusions

Patterns are – for Dennett – compressions of data or epistemic shortcuts. Patterns are observer-dependent, both from the perspective of the individual observer as well as the type of observer – what sort of creature it is. Multiple patterns may thought of as sitting on the same underlying data, their various appearances are due to distinctions in the observers, but simultaneously patterns may be thought of as being the result of indeterminate (or multiple) causally generative processes. Patterns are both perspective- and scale-dependent; but they are never-the-less real - they are there to be discerned by a capable observer, however great their complexity. The noisiness - and subsequently required idealization - means they are, though, mildly real.

¹⁶⁵ Dennett, 1991, pg. 45
¹⁶⁶ Dennett, 1991, pg. 43
¹⁶⁷ Dennett, 1991, pg. 44
¹⁶⁸ Dennett, 1991, pg. 42
However, it must be said that – at least within the framework of “Real Patterns”, taken as a single paper and not within Dennett’s larger work as a whole – it cannot be said that such a view of patterns has been proven, it seems to me it has merely been shown to be possible. Are the patterns with which we are here concerned – patterns of intentional behaviour within a rational framework (of which instances of ‘knowledge’ is one), or even objects such as maple trees and tulips – very much like the contrived example of bar code, in which the noise level can simply be turned up until it has as much relevance within the signal as the larger squares? Is the noise with which we are concerned of such a character that it could be interpreted as a coherent part of a different signal, or is that just a result of the artificiality of the example? Are the patterns of behaviour that we observe and that fit into larger frameworks of rationality and predictability really of a type such that multiple interpretations of the pattern are possible, and is this to such a degree that the stronger version of Quine’s Principle of Indeterminacy of Translation holds, as Dennett claims it does? These questions are not sufficiently addressed in this portion of Dennett’s work.

It also well established that there is a unifying and underlying causal structure to many of the types that are the referents of our concepts – tulips, for example, have a specific genetic code that makes them tulips – and this causal structure would place severe limits on how much the noise can be ‘turned up’. Bar code was constructed such that Dennett’s point about indeterminacy holds, just as the Life example is constructed such that the elements of the Turing machine that runs the chess program are – by his definition – “far from perfectly rational” since Dennett has again introduced an element of noise. Indeed, the noise that Dennett introduces is – by definition – random, and therefore by his own admission is not a candidate for compression, and therefore could not be a signal. Dennett relies too much on random noise as a lynchpin of his indeterminacy argument. It could be argued – contra Dennett - that the Life example with noise is simply not a Turing machine at all. Turing machines are by definition, digital devices that do not err due to noise or incomplete/imperfect elements.

My own view is that Dennett is right about three key points. First, pattern instances vary one from another (even as examples of the same type), and do so in – pardon the lack of rigour here – messy and varied ways. Trees can blow in the wind, be tall or short, standing or burning, plastic or carbon, pruned to a bonsai or as bushy as the untended willow in my parents back yard: they are all trees, and the way in which they vary one from another is – noisy. Whether ‘noise’ in this case has a random component in the perturbation from some paradigmatic case or prototype, or whether it is entirely non-random variations on those themes, does not matter: the point is that tokens vary.

169 Dennett, 1991, pg. 41
considerably from each other and are classified as a type according to the interests of the observer. Second, I also take it that Dennett is right that one person’s noise can be another’s pattern - *as long as that ‘noise’ is not Dennett’s genuinely random element* - but rather, an element of the signal that merely *appears* random to one observer. One person’s noise cannot be a compressible signal to another person – a pattern - if it is genuinely random. Thus, it seems correct to say that there are a multiplicity of possible interpretations of some chunk of the world. So much so for objects, so much more so for intentional patterns of behaviour. Lastly, I also take Dennett to be right in identifying the mind – or total epistemic state – of the observer as being what enables one pattern of the multiplicity of possible patterns to emerge as relevant to that observer.

That said, Dennett has not been clear about whether or not some intentional pattern is *there*, discernable-in-principle, or whether such a pattern is only there in the sense that it can be discerned by an external observer adopting the appropriate stance. The relationship between the complexity of the pattern and the role of the mind in observing or picking out that pattern has not been sufficiently explored, although I take it as firmly established that minds are involved at least to the extent that it (along with the rest of the observing system) need be present for the pattern to be privileged in some way (*noticed, picked out, or identified* in some way).

### 2.2.4.1 Support for Ontological and Epistemic Continuity

One reason I started with Dennett is because his take on patterns makes a *very explicit link* between simple, visual patterns of dots on the page and the (much) more complex patterns involved in folk psychological intentional concepts. Here is an explicit endorsement of each my claims regarding *ontological* and *epistemic continuity* made at the outset. Recall that *ontological continuity* is the claim that the patterns associated with the visual objects of perception (like shrubs) exist on a continuum with the more abstract patterns associated with intentional items (like knowledge). Recall that *epistemic continuity* is the claim that perception and conception also exist on a kind of continuum.

Dennett is using his two visual examples to motivate his view concerning the reality or ontological status of intentional states. The simple patterns could be seen as merely a kind of conceptual prodding, but they are presented by Dennett as being much more than that. They represent specific, concrete examples of how we can interpret patterns in various ways, and since they are “much
simpler, more readily visualized, and uncontroversial sort of pattern[s] the lessons drawn are correspondingly harder to deny. These visual patterns are not mere analogies, but are meant to be regarded as patterns occupying the simple end of an ontological spectrum that leads in a direct way up to the more complex patterns associated with intentionality; this is precisely what I mean by ontological continuity. Dennett may be offering an account of degrees of realism associated with this continuity, but continuous he certainly means it to be, and he may be particularly concerned with the ontological status of intentional states, but as Haugeland says, “... the issue is not intentionality at all, but rather being. ... all the main points are made in a more general way, in terms of patterns. Intentional states are just a special case ...”171. “Special case’, indeed; I believe this passage can be interpreted as support by Haugeland for ontological continuity as well.

Dennett also makes clear that patterns are not merely visual, but in adopting the right stance or perspective that bring the patterns out, we are performing an overtly intellectual act. Just as patterns on the page can be better seen by squinting, or turning the page this way and that, patterns that are to be perceived by adopting a particular stance can be brought to light by performing a kind of intellectual squinting or making a particular kind of intellectual effort. Finding the right stance by which to view some set of events is to be ...

“... lucky or clever enough to hit on the right perspective. [The patterns of chess moves in the Life world, see below] are not visual patterns but, one might say, intellectual patterns. ... The opportunity confronting the observer of such a Life world is analogous to the opportunity confronting the cryptographer staring at a new patch of cipher text, or the opportunity confronting a Martian, peering through a telescope a the Superbowl game. If the Martian hits on the intentional stance – or folk psychology – as the right level to look for pattern, shapes will readily emerge through the noise. [emphasis mine]”172

I take it that this is direct support for a brand of epistemic continuity; the kinds of cognitive activity associated with developing the concepts associated with adopting a ‘chess stance’ are of a kind with the cognitive activity involved in seeing a visual pattern on the page.

---

170 Dennett, 1991, pg. 31
171 Haugeland, 1998, pg. 267
172 Dennett, 1991, pp. 41-42
2.3 Haugeland; Context-Dependence and Normativity

Haugeland argues that Dennett equivocates between two levels of pattern, the level at which elements are defined as *elements of patterns* and the level of the element itself. To illustrate: there are two levels at which one might talk of ‘real patterns’ in the Life example – there is the individual pattern itself (say a glider, or set of gliders) and there is a larger context or higher-order pattern (the game of Life as Turing machine) in which that individual pattern may be a participant, and which makes that element that particular object (the gliders *as* components of a Turing machine, a memory state or part of the computing architecture). Haugeland is concerned with the role that patterns (elements) play within larger patterns (contexts) and how that role endows the patterns (both elements and context) with their status as objects. Here, a notion of context dependence is developed in which the elements and the context are mutually defined in a kind of virtuous circle. I shall call this sort of context dependence *strong context dependence*.

Being an object is to be *taken as* that object (rather than another object or rather than some random agglomeration of Ontology I), and it is the context of pattern-within-pattern that generates normativity, for Haugeland. Normativity is intimately involved in recognition, for to recognize a thing is to recognize that thing *as something* (a tulip or tree in the garden, an inspired pick-and-roll on the basketball court). I shall here concentrate on the ontology of pattern, and of particular concern is the notion of strong context dependence and its relation to normativity or *objecthood*.

Haugeland also wants to buttress the idea of a ‘stance’ with the notion of ‘commitment’, and it is the kind of interplay that exists between levels of pattern that demands this stronger notion. The constitutive standards of rationality, viewed as an over-riding contextual pattern into which intentional acts fit, is taken as such a commitment. The commitment is an epistemic act, it is something brought to a situation by an observer, and it enables certain patterns to emerge. The commitment is an operational notion, it is a cognitive *act*, and a question is raised as to the ontological status of the patterns that emerge: if they require an operational component, are they *really there*, independent of the observer making the commitment to see them?

A point I shall start to bring out here is the idea that such patterns may indeed be regarded as being discernable-in-principle – they can be regarded as *really there* - even though they necessitate an operational component in their identification conditions. Hence, I want to be able to say that the constitutive standards of rationality are discernable-in-principle, as are the intentional patterns that
emerge (an ontological claim, according to Haugeland), yet what is required to identify those standards and those patterns is a commitment to, or participation in, those standards and patterns themselves (a epistemic act). Discernability-in-principle, then, is argued to be independent of the question of an operational definition: some patterns may only be defined as being the sorts of things that ‘you know one when you see it’, but that does not mean that pattern collapses into a sort of *vacuous* operational notion, as long as ‘discernable’ is given a broad enough toolkit to specify the relevant identification conditions.

### 2.3.1 Strong Context-Dependence

In the bar code example, the bits only count *as bits* in a signal (rather than just dots on a page) if the series of bits is itself considered as a signal (or pattern). The ontological status of the dots is not at issue - there they are on the page. What is at issue is whether the dots have the status of bits; “A particular … black dot on white paper would not *count as* a bit or pixel except insofar as it is a component in some relevant mathematical or visual pattern.”

Extending that idea, “each glider is the glider it is whether or not it also counts as (say) a token on the tape of that Turing machine.” The status of these individual patterns *as patterns* depends on their participating in “higher-order patterns or structures”, in which “they can be relied upon as *components*”. So a structure that is identified as a pattern can simultaneously be both a pattern and a component of a pattern.

A structure is a candidate for pattern-hood if it is *special* in some way, for there needs to be a “motivation for picking out this particular sequence [over another]”. There are thus two ways in which some structure may be special. It may be that a sequence exhibits some interesting stability on its own, without reference to a higher-order pattern, for instance a glider; this Haugeland calls specialness “*from below*”. The glider may be a component of a Turing machine, however, and it is that component by virtue of participating as an element in a larger pattern; this Haugeland calls “*specialness from above* – something’s being noteworthy *as a pattern* not by virtue of how it’s built out of elements, but by virtue of how it participates in or contributes to something else.”

---

173 Haugeland, 1998, pg. 269  
174 Haugeland, 1998, pg. 270  
175 Haugeland, 1998, pg. 271  
176 Haugeland, 1998, pg. 271  
177 Haugeland, 1998, pg. 271  
178 Haugeland, 1998, pg. 270  
179 Haugeland, 1998, pg. 271
The higher-order pattern in which another pattern or object is a component can be regarded as the *context* which provides a background by which the object can be taken as that object – the context, then, plays a normative role since it contributes to the identification conditions of the object. Thus, object-hood is context-dependent. Elements of a pattern inherit their object-hood by virtue of participating in a pattern, and that pattern is a pattern by virtue of being made up of those elements. Context-dependence here is a two-way street, and - more than that - the two levels mutually define each other in a holistic manner in which independent identification of each level - outside of a bare recognition of each, defined as a purely *operation* notion - is nonsensical. This slightly circular notion shall now be made clear.

Haugeland points out Dennett’s use of an operational definition of a pattern (that it is something *recognized* rather than defined) allows for a notion of context dependence that is quite holistic – and in this point follows exactly the line I shall bring out in Dreyfus. Context dependence is:

“… frequently understood on a broadly inferential model:

1. Any instance of I, in context C, would be (or count as) an R.
2. Here is an instance of I; and it is in context C.
3. So, here is an R.

This presumes, however, that C and I are identifiable as such independently, and that the recognition of R is then just drawing a conclusion – not really a *recognition* at all.”

Haugeland does not think that contexts and instances can be nicely separated like this - I shall bring this point out in detail with Dreyfus - and I note for now that the problem originates in the supposition that the elements can be independently defined as elements, or if they are to be bound up in the context in this manner, then that context can itself be taken as an independently-defined element. Haugeland makes many of the same points, and applies them to Dennett’s view of pattern:

“First, many relevant patterns .. do not seem to be made up of well-defined bits or elements. … Second, the account of patterns as orderly arrangements of predetermined elements is an invitation to metaphysical reduction … [which] runs counter to Dennett’s motivating insight that ‘real patterns’ might be of distinctive ontological status and interest. Third, [the higher-order patterns’] recognizability
would have to be *via* prior recognition of those elements; and that would be a version of epistemological foundationalism.\(^{180}\)

Dennett has claimed, emphasized by his nomenclature of *mild realism*, that the intermediate regularities revealed by the intentional stance are patterns of some interesting ontological status. The reductionist implication Haugeland is pointing out would seem to mitigate any special ontological claims. The epistemic foundationalism follows, in that the elements to which the larger patterns are reduced must have a prior epistemic determination in order to be used to generated the recognition of the larger pattern of which they are elements. Here, we bump up against Amenability to Algorithmic Description (AAD) – the idea that a pattern can be viewed as ‘computed’ from independently-defined elements – and this shall be analyzed in depth in the next section. Here, I note the role that strong-context dependence can play: it brings out a kind of *recursive* or *circular* complexity. For Haugeland, the circularity is not so much vicious as it is virtuous;

“… the two notions of pattern join forces … In this larger conception, the ‘elements’ of an orderly arrangement need no longer be thought of as *simple* (‘elementary’), like bits or pixels, or even as independently identifiable. On the contrary, they might be quite elaborate, elusive and/or subtle – so long as some relevant creatures are …able to recognize them. This recognizability, in turn, can perfectly well depend, in part, on their participation in the arrangement (=the context) of which they are the elements.”\(^{181}\)

So here we have what I shall call *strong context-dependence*, in which patterns gain object-hood by virtue of the higher-order patterns in which they participate, and those higher-order pattern gain *their* object-hood, in part, by virtue of the sub-patterns of which they are composed. Haugeland regards this circularity as virtuous since each level of pattern can be regarded as contributing to the other level’s identity conditions. As we shall see in more detail when we come to analyzing AAD (Dreyfus, Section 2.4), this circle may be regarded as virtuous only if a notion of bare recognition is maintained; prior to that analysis it is entirely reasonable to assume that the identity of either level of pattern is not independently defined outside of an operational notion of recognition. Haugeland seems amenable to this sort of claim, given the ‘so long as some relevant creatures are … able to recognize them’ in the longer quote above. So, despite the fact that such recognition depends ‘in part’ on their ‘participation in the arrangement’, the status of groups of elements as patterns

\(^{180}\) Haugeland, 1998, pg. 275

\(^{181}\) Haugeland, 1998, pg. 275
remains dependent on a creature’s participation as recognizer. In other words, discernability-in-principle (a pattern’s identification by virtue of its participation in a larger pattern) does not preclude the necessity of recognition as an operational component of the identity conditions.

2.3.2 Normativity, Objecthood and Discernability

As stated earlier, the notion of objecthood – to be taken as one particular sort of object rather than another – is intimately related to the notion of normativity, for to take something as an object is to be open to error; one can make a mistake in such an attribution. Here, I am concerned with how the issue of normativity is related to the context-dependency of patterns. For Haugeland, it is by virtue of the fact that we have these “‘patterns’ at two different levels … in such a way that … each requires the other …”\(^{182}\) that allows for the normativity of recognition to be made precise. Strong context-dependence, then, is what provides for a notion of normativity within the metaphysics of pattern.

The possibility of mistake is at the heart of recognition, for, “To recognize something is to respond to it in a way that distinguishes it from other things; to recognize it is to tell apart.”\(^{183}\). Recognition is not mere differential response, though, since mere differential response does not invoke the possibility of getting it wrong. More specifically, there must be some “one single factor in the causal background of the response [that] has … a special status and importance ... and then, if that single factor could be identified as its object, the response would amount to a recognition”\(^{184}\).

Recall that an aggregation of bits may count as a pattern in two ways. First, given a formal definition of pattern it may be a candidate for signal compression all on it’s own (specialness from below), and second, it may play some role as a component within a larger pattern (specialness from above). A chess piece is a paradigmatic example. Here, the larger pattern (the rules of chess) is what makes some participating object a chess piece, and not merely a horsey-shaped bit of plastic. It is a ‘horsey’ object by virtue of its specialness from below (it looks like a horse), but it is a ‘knight’ by virtue of its specialness from above (it moves in certain ways on the board). In the case of chess, which is an artificially clean and clear (really - digital) example of a higher-order pattern, the rules are understood by all (and can usually be codified in some way); the knight-ness of the object is fully constituted by its role within the larger pattern of chess. But of course, a knight does

\(^{182}\) Haugeland, 1998, pp. 275-6
\(^{183}\) Haugeland, 1998, pg. 272
not have to be a horsy-shaped object – it may be a cup, a saucer or my mother’s left shoe – as long as all the players agree. What makes the knight a knight then, is its role within the larger pattern. Yet – and this is the bit that provides for strong context-dependence – it is the movements of the individual pieces that enable one to call the game chess in the first place.

Thus, the pieces are what they are by virtue of the game, but the game is what it is by virtue of the pieces. It is, for Haugeland, the “taking on a commitment to the legality of the game” and commensurately a “commitment to recognize the pieces, positions and moves – and recognize them correctly” that enables the introduction of norms that govern the game and its pieces. Given the ability to recognize patterns, the interplay of higher- and lower-order patterns provides for a means by which both levels of patterns may be provided with object-hood. More generally, “if a larger arrangement pattern is constitutive for the domain of its elements ... it can induce the norms by which those elements can themselves be recognition patterns. That is, the elements can be criterial for the correctness of their own recognition, and in that sense be objects.”

Note here that recognition is assumed to be a skill that is brought to the table (sic), it is not reduced to a formal notion of independent definition; recognizing is that you quite simply “know one when you see one”. What is being recognized is not defined prior to participating in the game, nor – in either my view or Haugeland’s – can it; “....recognition patterns need have no separate identifiability at all (as patterns of anything else, for example) apart from their recognizability in context.” Dreyfus shall provide further evidence to this point.

Chess, of course, is an artificially simple example compared to intentional states and the constitutive role that rationality plays in defining or constraining those states (recall the role of noise in Dennett’s examples). In chess, there is no noise, all of the interesting complexity inherent in strong context dependence can be found within the limited and artificial world of chess; “for digital games are quite exceptional in the explicit precision with which their constitutive standards can be articulated ... Even physics ... is less well spelled out in its ontological foundations than is chess, and more tolerant of anomalies in its concrete practice.” The complexity that arises from a chess example can be presumed to be of a lesser order than that of intentional items. These lessons can, in the spirit of ontological continuity, be applied to much larger picture; “The inner coherence

---

184 Haugeland, 1998, pg. 276
185 Haugeland, 1998, pg. 278
186 Haugeland, 1998, pg. 279
187 Haugeland, 1998, pg. 279
188 Haugeland, 1998, pg. 282
and genius of Dennett’s position lies not in the analogies among [patterns, games and the intentional] but the underlying unity that binds them all together, that makes them all distinctions among ways of being.”  

In the case of intentional items – such as knowledge – it is rationality that plays the role of a higher-order pattern that prescribes the constitutive standards of the domain. The point of the noise in earlier examples was to represent the kind of imperfect rationality that we exhibit (see Chapter 5 for arguments that this is indeed so) as well as the massive variance in behaviour that can be associated with a single intentional description (think of all the ways that ‘wanting a carrot’ might be made manifest in behavioural terms). With regard to the constitutive standards governing intentional behaviour, then, “there need be no articulate specification of these standards: all that is really essential is an ability to tell, in any given case, whether the standards are being met.”  

In adopting the intentional stance, we are placing patterns of behaviour inside a larger pattern or context, in which we are capable – as a matter of practice – to make sense of, or identify, the individual intentional items. We are left with, again, an operational notion of pattern defined in terms of the act of recognition.

The role that rationality plays in providing the constitutive standards (the higher-order pattern) by which individual patterns of behaviour are bestowed with intentionality is crucial, for Haugeland. It is more than simply adopting a stance – as Dennett would have it – but a matter of commitment. For Haugeland, “[a]dopting a stance is taking a stand.”  

Why the distinction? Because it is only by committing oneself to constitutive standards – standards which one must insist upon on pain of “giving up the game” – that one is in a position to recognize activities as being of some determinate type. The point is that the intentional stance is not a sort of vantage point by which one can suddenly see all the intentionality, it is a commitment to try and make sense of things; “Only against a genuine and determinate possibility of failure can there be any meaningful success.”

By this Haugeland means that rationality itself is a pattern by which one makes sense of other intentional patterns, and it too, requires recognition. That recognition itself requires an insistence on correctness; “In other words, for intentional attribution, we insist upon rationality: in confronting apparent breaches of the standard, we first attempt to rectify them … and, failing that, we give it

---

190 Haugeland, 1998, pg. 285
190 Haugeland, 1998, pg. 283
191 Haugeland, 1998, pg. 285
192 Haugeland, 1998, pg. 284
193 Haugeland, 1998, pg. 284
194 Haugeland, 1998, pg. 284
up.” Hence, one is committed to a constitutive standard, one does not just take the stance of a constitutive standard. The constitutive standard of rationality is a sort of global pattern to which we commit ourselves, and provides a context within which the intentional patterns emerge to view. In that sense, the constitutive standard of rationality can be thought of as more than merely operational – it is what we commit ourselves to in order to enable the recognition of the more operationally-flavoured intentional patterns subsumed by it. That is the distinction of commitment versus stance, for Haugeland: the constitutive standards to which you commit yourself are treated as more then merely operational. I do not entirely agree with what distinction entails, as I argue below, but I do agree that the distinction allows for the conceptual priority of the constitutive standards over the sub-patterns subsumed under those standards.

2.3.3 Concluding Remarks: Are Constitutive Standards Recognition-Independent?

So here we have a view of patterns in which there is an interplay between individual patterns and the higher-order patterns of which they are a component. Strong context-dependence of patterns within a higher-order pattern imply a kind of mutual dependence and mutual definition. This interplay leads directly to a concrete notion of normativity; given that one makes a commitment to the constitutive standards, one makes a commitment that the sub-patterns will make sense. For Haugeland, the requirement that the commitment comes first, and the recognition of sub-patterns second, indicates that the constitutive standards of rationality have a status that is independent of our perception. The constitutive standards to which we commit have a status that renders them available to us in a way that is independent of acts of perception.

Although I recognize the temporal priority of the commitment over the act of recognition (first I commit to make sense, then I see the intentional patterns), I do not agree that the constitutive standards gain any sort of special status over the act of recognition other than a practical one (one can’t do one thing before doing the other). This notion of normativity does not do away with the need for an operational notion of pattern, which in turn depends upon the prior existence of recognition as a kind of independent skill that is brought to bear on the situation. Constitutive standards also require recognition. It is thus on my commitment to the mutually-reinforcing existence of recognition and constitutive standards (and not just patterns) that I part ways with Haugeland. For Haugeland, the constitutive standards of rationality have a logical priority and a

---

194 Haugeland, 1998, pg. 284
status independent of perception, but for me both notions can be thought of as being *simultaneously* in play - even though the constitutive standards must be committed to prior to recognition of the intentional sub-patterns.

I would argue that the commitment to rationality is either a choice or is heavily dependent on my perceptual systems. Either way, I remain neutral on any further claims about the special status of the constitutive standards of rationality, or cogency, as independent of acts of recognition. On my view the constitutive standards must themselves be recognized in order to be chosen – they do not sit ‘out there’ with a distinct status that delivers them to me independent of the act of recognizing them. To the extent that I *seem* unable to choose irrationality, to the extent that it is difficult to live that way, then to that extent my choice is limited by my perceptual systems and my brain. Hence, those constitutive standards are again rendered perception- or recognition-dependent. To the extent that I can choose to be irrational from the perspective of my peers, then to that extent I can choose to recognize a different constitutive standard – and perhaps recognize different sorts of intentional patterns as a result. That is not a very useful choice, to be sure, and would certainly exclude me from meaningful participation in the socio-linguistic community that will later emerge as so important. The distinction of my view from Haugeland’s can be viewed as relatively minor if we were not concerned with the recognition-independent status of the constitutive standards, but is a major difference if that property of the constitutive standards of rationality is at issue.

One major point of agreement between myself and Haugeland, though, has emerged as a question of priority. If we cannot reduce a pattern to some definition independent of recognition, and if recognition is to be regarded as a kind of independent normative skill (the skill being dependent on successful recognition of the pattern) then what appears to be some sort of conceptual collapse - pattern requires recognition, and recognition requires pattern – is really just the simultaneous operation of two independent notions of pattern. The independent existence of patterns – that they are discernable-in-principle – can be (conceptually, if not practically) separated from the necessity of an operational component of recognition.

Unlike Haugeland, though, I do not draw a distinction between the constitutive standard of rationality, and other higher-order patterns, other than to acknowledge those ‘very high’ order patterns are more deeply entrenched in our ways of living and our perceptual systems themselves than other sorts of higher-order patterns that play constitutive roles (like the rules of chess). It is precisely this vicious circle that I shall argue C-nets embedded in a socio-linguistic community break, and so the details of that account shall wait until Chapter 5. Thus, the answer to this central
question - are the patterns discernable-in-principle, or are they purely operational in character? - can, on my view, be again posed with a view to the status of the constitutive standards themselves. The line I am bringing out as my disagreement with Haugeland is that these two answers are not mutually exclusive for the constitutive standard of rationality, just as they are not for other patterns. The constitutive standards may be discernable-in-principle, like any higher-order pattern (ie, out there in the world, waiting to be discerned, as it were) but the discernment may still be purely operational in character, in that there must be some actual and present cognizing system to play a part in the discernment.

A main point of agreement, however, has to do with the nature of ‘discernment’, and what tools one allows oneself to use in providing a means of discernment. The over-riding argument I am making is that some patterns are not discernable in some restricted languages: intentional patterns are not discernable in the reductive language of traditional computing/physicalist models, for example, or on a simple non-recursive view of patterns in which large-scale patterns are made up of independently defined elements. Given those restricted tools, some patterns of interest are not discernable-in-principle. Given a larger toolkit of means of identification – operational C-Nets, for example – then those same patterns are discernable-in-principle. I deny that the necessity of an operational element of recognition leads to a collapse of the discernment-in-principle character of pattern. What I do advocate is that the discernment-in-principle notion be broad enough to include the participation of a relevant observing creature in its identification conditions. Hence, I do not see these two notions – discernment-in-principle and the necessity of an operational character in the conditions of that discernment – to be in competition.
2.4 Dreyfus; The Kinds of Patterns We Recognize

Dreyfus sees pattern recognition as an essential skill that underpins and functionally binds together a number of distinct cognitive tasks, those tasks being (roughly) categorized as game playing, natural language understanding or translation and general problem solving. Each of these tasks “presupposes success in the field of pattern recognition”\(^{196}\) and pattern recognition is really Dreyfus’ core concern. How we recognize patterns is the topic of the next chapter, here we are interested in what a pattern is. What I want to extract from Dreyfus’ work is a characterization of a pattern’s Amenability to Algorithmic Description (AAD). What I mean by AAD shall be made explicit in the next section, and here I am laying some foundations by which to generate that analysis; I am now concerned with the question as to whether patterns can be regarded as composed of discrete bits, traits or properties by which acts of classification may be performed.

One of the points Dreyfus drives home is that pattern recognition is not always just a matter of classification, and that when states of affairs are recognized as being an instantiation of a particular pattern, there may be no corresponding set of defining traits; “In some cases … [the explication of the traits defining a member of a class] … is not even possible. … To appreciate this point we must first get over the idea … that pattern recognition can always be understood as a sort of classification.”\(^{197}\) Here, Dreyfus is concerned – not just with the possibility that patterns may be so complicated that “so many traits are needed for discrimination [that] the problem of exponential growth threatens”\(^{198}\), which is a practical or pragmatic affair – but with an in principle impossibility of trait-by-trait pattern identification. Again, patterns are not to be thought of as being composed of predetermined bits – however many and however complicatedly related – but as being of a different order of complexity. Here, we are reminded of Dennett’s intermediate regularities that exhibit a kind of indeterminacy, and – as we shall see – this is also related to what I have termed strong context dependence in Haugeland’s patterns-within-patterns.

2.4.1 Three Kinds of Patterns

Dreyfus argues that in making the assumption that pattern recognition is a kind of classification “three distinct kinds of pattern recognition are lumped together”\(^{199}\). In each sort of recognition it is the use of context to reduce ambiguity that is a key feature, and it is in bringing out the distinctions between these kinds of recognition that allows the metaphysics of pattern to emerge. In no case is

---

\(^{196}\) Dreyfus, 1999, pg. 97

\(^{197}\) Dreyfus, 1999, pg. 123

\(^{198}\) Dreyfus, 1999, pg. 121
the pattern to be considered as a concretely-defined list of features, or independently-defined sub-patterns.

The first sort of recognition is what Dreyfus (following Gurwitsch) calls "the generic," which is a simple kind of object recognition in which the object is recognized as being of a certain type, such as "the recognition of a certain object as a pencil." Here, "it is our purposes [which] serve to select which features are significant, and, among these, certain features which are crucial." The point here is that – following Wittgenstein – Dreyfus emphasizes that the question as to which features are significant and which crucial is an open-ended one, and it "changes with our changing purposes and knowledge." Perhaps it is crucial that it makes a mark (is a dirty stick a pencil?), or has a point (does a large child’s pencil have a ‘point’?). In reducing the ambiguity of what counts as a particular object, the context that is available to do so is open-ended, since the background knowledge and sense of purpose we may bring to bear is open-ended. Hence, even this simple kind of pattern – an everyday object – is not a clear candidate for context-free element identification. Artifacts are clearly the objects at issue here, or things that may serve a particular purpose.

The second sort is "the recognition of resemblance," in which particular patterns may be independently recognized, but only after the context has been allowed to play a "determining role." Here, an emphasis identical to Haugeland’s strong context dependence is at play, in which there is a reciprocal interplay between the pattern and the larger pattern in which it plays a role. Dreyfus’ use of the word ‘resemblance’ here is somewhat unfortunate, since resemblance plays a role in all three types of recognition. Here, though, the classification of this type of recognition hinges on a denial of there being isolable traits – so perhaps a better nomenclature might be ‘contextual recognition’.

The best example here – and it is directly analogous for Dreyfus to the case of linguistic disambiguation – is a human face, and the role that the individual features of the face and the face

---

199 Dreyfus, 1999, pg. 123
200 Dreyfus, 1999, pg. 123
201 Dreyfus, 1999, pg. 123
202 Dreyfus, 1999, pg. 124
203 Dreyfus, 1999, pg. 123
204 Dreyfus, 1999, pg. 124
205 Dreyfus, 1999, pg. 124
206 In terms of linguistic disambiguation, it is the not just the surrounding words that disambiguate meaning, but also the actual surrounding context and shared understanding of those communicating. Dreyfus gives a wonderful example of “Stay near to me!” and the different meanings ‘near’ might take on in the case of two space-ship
itself play in mutually determining the expression. The expression of a person’s eyes contribute to
the overall expression that is interpreted on the face (this seems obvious), but likewise the
expression of the eyes themselves may be dependent on a more global feature of the facial
expression; “the expression of a person’s eyes, for example, may depend on the whole face in such
a way as to be unrecognizable if viewed through a slit.” On a larger scale, a face – with a
neutral expression on it – will be interpreted (or recognized) in different ways if that face is
surrounded by different contexts. The example Dreyfus cites is an experiment by Pudovkin in
which the same (quite neutral) face was shown in three different contexts (looking at a dish, a dead
woman and a child playing with a teddy-bear) and the expression was seen as pensive, sad and
happy respectively. The point here is that “In this sort of resemblance, the notion of recognizing
the pattern in terms of isolable traits make no sense” - and I take the lesson to be the point that the
pattern itself is likewise not composed of those isolable traits.

At this point I take an example from Hofstadter which I believe makes much the same point but
with ‘context’ extended to include the knowledge and background of the viewer. An old black and
white film is on television, watched by a fly, a child and an adult. The film is sad, and there is a
moment of great pathos and irony in a situation that takes place in a kitchen while someone eats a
bowl of soup. The fly sees, presumably, black and white dots. The child sees someone eating a
bowl of soup. The adult sees a sad and ironic moment, full of purport and meaning surprisingly
analogous to their own life. Which pattern is on the television? Surely all these, and more, and so I
take it that the pattern itself is highly dependent on the context - to the extent that it makes little
sense to speak of the pattern independently of the context (which in this case includes the state of
mind of the viewer) – and the context is also affected by the pattern (the viewer responds to the film
in such a way that allows for the pathos to be brought out – isn’t this just what a good director
does, build a reciprocity between film and audience?). I would argue that each of these patterns –
that perceived by the fly (dots on the screen), child (the eating of a bowl of soup) and adult (the
extended sense of irony) – are all discernable-in-principle, yet each are simultaneously dependent
on an act of recognition. That act of recognition is what defines the available context (eyeballs only,
a child’s perspective or an adult’s long and interesting history with the world) by which the pattern
emerges. The ‘discernibility’ criteria can be, I argue, extended by virtue of the cognitive tools the
commanders exploring deep space, and a man and woman holding hands and about to enter a dark alley. ‘Near’ does
not have a meaning independent of a great deal of contextual information that extends far beyond the page or
conversation.

---

207 Dreyfus, 1999, pg. 125
208 As quoted in Merleau-Ponty, 1964, pg. 54
209 Hofstadter, 2007, pp. 152 to 154
observer brings to bear. That is really just another way of saying that the discernability criteria is open to be defined in a way that depends on what parts of the universe the observer can bring into the situation as context. Discernability-in-principle operates alongside - in conjunction and simultaneously with - the cognitive tools of the observer.

So the sort of context-dependency talked about by Haugeland, and the multiple perspectives talked about by Dennett, each lead to the notion that an appropriate context to think of in which a pattern is embedded – and from which it derives its objecthood – may include an extraordinary amount of material, including (but not limited to) the brain states and personal histories of the observers.

Lastly, Dreyfus talks about a sort of pattern recognition which he calls “the recognition of similarity” in which “there may be no common traits, even overlapping ones …” “[or even] context-dependent ones”. This kind of recognition is tightly bound up with Wittgenstein’s notion of family resemblance. Family resemblances differ from class membership in that family resemblances “are recognized only in terms of real or imaginary examples … [they allow] a spectrum ranging from the typical to the atypical … this sort of recognition … is accomplished not by a list of traits, but by seeing the case in question in terms of its proximity to a paradigm (ie typical) case, such recognition gives us another kind of openness and flexibility.”

Here, concepts like ‘grace’ or - I shall later claim – knowledge, are typical examples; these “can not be defined in terms of necessary and sufficient conditions” but are rather closer or farther from typical cases, and such cases may not be actual but even imaginary.

I shall return to what ‘recognition’ means in these cases, particularly what sorts of cognitive skills one needs to bear in order to be successful, but the point being made here is to deny that there are definite lists of traits – or elements of the pattern – that can be independently defined as constituting such patterns. This notion of family resemblance is, for Dreyfus, a matter of a “network of crisscrossing similarities” in which no two family members “need have any identical features for them to share a family resemblance”. There is no disjunction of common properties, and Dreyfus emphasizes that formalizing the notion of family resemblance in this way “would eliminate a kind of openness to new cases which is the most striking feature of this form of

---

210 Dreyfus, 1999, pg. 126
211 Dreyfus, 1999, pg. 126
212 Dreyfus, 1999, pg. 126
213 Dreyfus, 1999, pg. 126
214 Dreyfus, 1999, pg. 126
215 Dreyfus, 1999, pg. 127
216 Dreyfus, 1999, pg. 126
A new example of ‘graceful’ may be noted, belonging to the family of resemblance that is graceful, the traits of which are somewhat similar to some given members, but “without being exactly similar to the traits of any of them”.

2.4.2 Facts, Patterns and Language; The Ontological Assumption

What Dreyfus is really arguing here is that the human world – our manifest image (following Sellars) – cannot be understood in terms of discrete elements that combine to form patterns or concepts. That the world can be understood this way is what he calls the ontological assumption; “The ontological assumption that everything essential to intelligent behaviour must in principle be understandable in terms of a set of determinate independent elements … is only an hypothesis [that] reflects two thousand years of philosophical tradition reinforced by a misinterpretation of the success of the physical sciences”. This is a part of a long tradition, from Plato to Leibniz and “reaches its fullest expression in Wittgenstein’s Tractatus, where the world is defined in terms of a set of atomic facts which can be expressed in logically independent propositions”.

It has to be admitted that a great deal of success has been reached, in the physical sciences in particular, by thinking of the world as a great soup of Ontology I, a great “set of independent interacting elements” - but this success does not extend any lessons regarding the validity of the ontological assumption when it comes to the human world. One quickly encounters a vicious infinity of regress when one attempts to do so, since the sort of context one brings to bear on a given situation and plays a role in the identification conditions of the particular pattern that is being picked out of the multiplicity of patterns available for discernment.

This brings to mind Hofstaedter’s fly, for whom the world of the television is (presumably) only a collection of independent bits, the black and white pixels. But the adult has a much more nuanced view of the film, and to bring out the patterns they see, one needs to take into account their very human situation which is not formalizable and is open-ended in interesting ways; the lives people have lived, the choices they’ve made - in short their entire worldview is relevant. It is “in our sense of the situation which enables us to select from the potential infinity of facts the immediately

Dreyfus, 1999, pg. 126
Dreyfus, 1999, pg. 126
Dreyfus, 1999, pg. 207
Dreyfus, 1999, pg. 211
Dreyfus, 1999, pg. 213
relevant ones, and once the relevant facts are found, enables us to estimate their significance.”\textsuperscript{222} Here, one is reminded of Dennett’s multiple perspectives, each of which extracts a particular fact from the swirling background of Ontology I. Which collection of dots, combined with which occurrent visual patterns in the remainder of the room, combined with what sets of neural patterns in the minds of the viewer that store past experiences, etc. etc. is it that makes the scene on the television \textit{poignant}? Give that fly access (and the means to program) an infinitely powerful computer, and still it could not answer. To answer that question, to pick out the relevant patterns, one must live a human life – one must \textit{participate in} our shared culture and life.

The context-dependence of elements forms a kind of infinite regress, for Dreyfus, for if elements are to be given relevance (from an infinite background of potentially relevant facts) by reference to a context, that context too must be recognized in its own right, and “the number of facts that might be relevant to recognizing a context is infinite too”\textsuperscript{223} One imagines an appeal to a higher context to fix the relevance of that context, and so on; unless some elements are fixed and intrinsically relevant, one who views the world as made up of discrete, independent elements forming larger and larger ones “will be faced with an infinite regress of contexts.”\textsuperscript{224} There is an ultimate context for us which is simply \textit{the life we live}; “What has to be accepted, the given, is – so one could say – \textit{forms of life}.”\textsuperscript{225} I shall have more to say about the nature of the infinite regress in Chapter 3, and leave a more detailed articulation of the argument until then.

But it does seem clear that understanding natural language is not like understanding a physically-described system. We bring to bear a shared understanding in the form of a set of background assumptions (our common-sense or world view) - open-ended and flexible in its own right - and we disambiguate meaning by reference to elements of global context, which is in turn reliant on the meaning of individual elements; “We seem to understand the situation in terms of the meaning of the words as much as we understand the meaning in terms of the situation.”\textsuperscript{226} The reciprocal nature of context and meaning is exactly analogous to Haugeland’s \textit{strong context-dependence}, and we manage to avoid Dennett’s \textit{strong indeterminacy} - not by upping the computational power applied to individual elements - but by using our human understanding, by applying our human pattern recognition systems, and buttressing these abilities with reference to the external framework of a shared language.

\textsuperscript{222} Dreyfus, 1999, pg. 218
\textsuperscript{223} Dreyfus, 1999, pg. 220
\textsuperscript{224} Dreyfus, 1999, pg. 221
\textsuperscript{225} Wittgenstein, 2000, pg. 226
\textsuperscript{226} Dreyfus, 1999, pg. 220
How it is that humans cut the Gordian knot and gain access to relevancy, similarity, and so on is the topic of the next two chapters. Neural networks – as a matter of empirical fact - display the sorts of responsiveness that get us out of this regress, but for now let us note that patterns are holistic in the sense that there is a kind of dependence on the context the viewer brings to bear, and that context is itself open-ended in interesting ways. Patterns cannot be regarded as simply being composed of atomic elements.
2.5 Patterns; A Hierarchy of Complexity

We have now a number of ways by which to describe the various degrees and sorts of complexity associated with patterns. I here present a number of axes by which complexity may be divided into types, and provide for a number of levels along each axis, corresponding to increasing degrees of complexity. The axes are: *amenability to algorithmic description*, *degree of mind-dependency*, and *degree of context dependency*.

Each axis will be divided into three sections, corresponding roughly to the three levels of complexity I wish to emphasize, and a set of examples are assigned to the appropriate level within each axis. A word of caution is required however; this analysis is not meant to be a logical analysis of complexity, nor is it meant to be an exhaustive account of the various means by which complexity can be measured, nor are the levels meant to be regarded as strict and definite dividing lines. Such a project is not just well beyond the scope of this thesis, but I do not believe that such an analytical project is even feasible – given the sorts of open-ended approaches to patterns that have emerged thus far. The hierarchy is meant to be a kind of thought-experiment, or a rough guideline, that is meant to generate intuitions about - and an understanding of - the kind of open-ended complexity displayed in the most interesting patterns, those associated with the top-most levels along each axis.

I shall take five examples as being typical of the kinds of patterns we are interested in; mathematical objects or shapes (a pyramid, sphere or ellipse), a natural kind (such as a tulip or maple tree), artifacts (such as desks and cars), a particularly human sort of pattern which I shall call a ‘gaming kind’ (a pick and roll in basketball, or a weak king side) and intentional items or events of uniquely human interest (knowledge, graceful or ‘believes that he can climb the tree’). I do not argue specifically for the inclusion/exclusion of each, but rely on arguments already made about the nature of patterns and concepts to make a considered judgement.

2.5.1 Amenability to Algorithmic Description

This axis is of particular interest since one of the lines of argument I am trying to build is that certain patterns cannot – in principle – be identified by traditional computing methods, which rely on algorithmic computation over isolable elements. By ‘computed in algorithmic fashion’ I mean that the relevant patterns can be defined, in an unambiguous way, by some computable function that...
acts over the independently defined or context-free elements of which it is made up. Clearly, there are two demands being made; first, that the elements themselves can be independently defined or taken as a kind of premise (call these representational primitives), and second, that some finite computational function in the form of some sort of algorithm can be found, the output of which is the classification (or recognition – note that these two terms can come apart in Dreyfus’ take on family resemblances) of the pattern.

That all concepts can be computed is, of course, a variant of the original Leibnizian project – for Leibniz all human reasoning and even skills “might be reduced to some sort of calculation”\textsuperscript{227} – this is a project that continued in the rationalist tradition, and culminates in its modern, formal variant as some interpretation of the applicability of the Church-Turing thesis.

A function $f(x)$ is considered to be Turing computable (or generally recursive) if “when some representation of the argument $x$ is put on the tape [of the Turing machine], the machine halts on a representation of the value $f(x)$”\textsuperscript{228} Here, $f(x)$ can be considered to be a classification function the output of which is some identifiable semantic item, and $x$ can be considered to be whatever representational primitives one wants (the visual dots that make up a digital image of a scene, the positions and trajectories of all the elements of Ontology I, whatever). Amenability to Algorithmic Description is meant to capture the feasibility of both defining each of $f(x)$ and $x$, and actually performing the required computation. In the extreme, $x$ could be considered to be all of the particles in the universe, $f(x)$ some function over those particles. The associated computation is – of course – practically impossible, given the physical limitations of the size of possible computing machines available; one would need vastly more particles to compute a set of particles $x$ than the number of particles being computed (ie to compute $x$ particles, one would need a number $y$ of particles to make up the computing device, such that $y >> x$; I take this to be obvious.)\textsuperscript{229} The task of identifying some limited $x$ is a deep problem, and central to the problem of computability; recall the fly and human watching the television screen. To identify the ‘poignant’ pattern, one must not only take into account some practically uncountable number of the neurons of the adults head (and make sense of them), but presumably some hard-to-define component of the the actual history of that person as well. I see no principled way of identifying $x$ (never mind $f(x)$...) for many of the patterns of interest.

\textsuperscript{227} Dreyfus, 1999, pg. 67
\textsuperscript{228} Blackburn, ed., 1996, Dictionary of Philosophy, Oxford University Press
\textsuperscript{229} The idea that all the particles of the universe are, in some way, computing themselves is an interesting diversion, but not relevant here. Computing here involves both representation in a syntactic domain and computation over that domain – see Section 4.2.3 for a discussion of computation.
Here, I want to distinguish between amenability in practice and amenability in principle, the distinction being one of whether or not the issue of mere computing power makes a difference. I want to make a principled distinction between those patterns whose functions could be both identified and computed, given infinite time and space in which computation can take place, and those patterns that could not, even given this open-ended possibility of computing power. Put more simply, if a function \( f(x) \) and domain \( x \) can be defined for a pattern, then such a pattern can be computed in principle, and - depending on the complexity of \( f(x) \) and \( x \) – may even be computed in practice. If \( f(x) \) and \( x \) cannot be defined for a pattern, then such a pattern simply cannot be computed – in principle. So amenability in practice means that conventionally available forms of computers could probably perform the necessary computations in finite time; amenability in principle means that – given some counter-factual universe in which infinite or near-infinite computing power is available (or perhaps quantum computing?) – the necessary massive or possibly infinite computations could be performed and the elements of the domain can be specified.

Note, then, that the distinction I really want to bring to bear has nothing to do with computing power, but rather the possibility of defining a relevant function \( f(x) \) and a domain \( x \) by which the pattern could be identified. So the in principle distinction is meant to be a metaphysical statement about the form of pattern, and in practice the distinction is more of an epistemic nature.

Here then, are the three levels associated with amenability to algorithmic description:

**Level I: Amenable to Algorithmic Description in practice:** Relevant patterns can be picked out by current-day computing standards, as the computable function \( f(x) \) and domain of elements \( x \) can each be identified, and the amount of computing required is on a scale available by current-day computing standards.

**Level II: Amenable to Algorithmic Description in principle:** Relevant patterns could not be picked out by current every-day computing standards, but it is more an issue of computing power than of the limits of computing itself. Both the computable function \( f(x) \) and domain of elements \( x \) can be identified, but their practical computability remains in question.

**Level III: Not Amenable to Algorithmic Description in principle.** No computer, however powerful, could pick out the relevant patterns as the patterns are not the sort of thing that is computable. One or both of either the computable function \( f(x) \) or the domain of elements \( x \) cannot be identified.
What I would like to draw attention to here is the limits of physical language in expressing the contents of our concepts. If one is to be a reductionist - a full-on, hard-core reductionist in the tradition of Weinberg, for example, or Churchland - who believes that the language of physics (what I call Ontology I) is sufficient to explain and predict all of the phenomena in which we are interested – then one would be committed to the position that all patterns are either Level I or II (Algorithmic) and the function $f(x)$ and domain $x$ for all patterns are expressible in the language of physics.

The examples I have outlined would be distributed along this axis as follows:

**Mathematical Objects:** Level I

**Natural Kinds** (maple trees and tulips): Level II, if genetic information can be solicited, possibly Level III if not since borderline cases may be problematic for Level II classification (is it a tree or shrub? A plant or animal? – humans are capable of just *deciding the matter* on some ad hoc pragmatic grounds, grounds which may not be thought of in advance as a part of $f(x)$).

**Artifacts:** Level III, given that many artifacts are functionally defined, and that function is relative to the life, interest and perspective of the user. Rocks can be chairs, if so used, and a car in a junk-yard becomes car-parts even though it may still look like a car.

**Gaming Kinds (weak king side):** Level II (given that ‘weak’ can be relativized to game statistics), Level III otherwise (*‘weak is something that would need to be sensed by the other player, and seems all-too-human and intentional in flavour to be computed).*

**Gaming Kinds (pick-and-roll):** Level III.

**Intentional Kinds:** Level III. Both the constitutive standards of rationality - the context in which individual actions may be regarded as intentional - and the individual patterns of behaviour within that context which are themselves intentional, are precisely the sorts of ‘intermediate regularities’ toward which we saw Dennett aim his intentional stance and about which he made his indeterminacy arguments. It seems to me that indeterminacy renders incomplete any view toward distinguishing particular patterns that relies on traditional computation. We also saw Dreyfus emphasize the need for an open-ended context to be brought to bear in disambiguating meaning and recognizing relevant similarities, and saw him undercut the whole notion that the world is composed
of isolable facts and determinate contexts. I have argued that to embue a television pattern with irony, pathos or some other intentional item, for example, a life of lived experience must be brought to bear. Intentional kinds are the sorts of things to which a person must bring their lived history to identify, and that lived history indicates that both \( x \) and \( f(x) \) are not defined in such a way as to be algorithmically computable – surely.

**2.5.2 Mind-Dependency**

I wish to distinguish between three sorts of mind-dependency, two of which we earlier saw Dennett merge into one sort – “These ‘design decisions … are incorporated into the design of our sense organs by genetic evolution and into our culture by cultural evolution … [and are responsible for] …what Wilfrid Sellars calls our *manifest image*”\(^{230}\). On the one hand we have a mind-dependency that results from the design of our sense organs, and our particularly human interests, and on the other hand we have a mind-dependency that results from our cultural immersion. Notwithstanding the fact that culture is – by definition – a group activity (language, a sense of discretion, etc.), it is learned in some way, and it is learned by each individual brain after birth.

I would like to further accentuate this distinction between evolved and acculturated mind-dependency by making a distinction between the mind-dependency resulting from what sort of creature we are and the mind-dependency resulting from the sort of learning (or training) we have each individually undertaken. Thus, I will define *weak mind-dependency* as being the result of being human, and it is thus associated with all humans, it something we all share – we all see red, for example\(^{231}\). I do not want to finesse this sort of mind-dependency with an argument about what sorts of patterns – precisely – only we humans can see (do chimps see social patterns such as altruism?) as that is not relevant to the point I am making. The point is that there is *some set* of patterns that you need a human-like mind to apprehend, and whatever set of patterns that comprises, are weakly mind-dependent. Some we may share, some we may not – that division does not concern me here. I will define *strong mind-dependency* as being the result of individual training, such as becoming a musician or an expert in chess, or modern art. *Strong-mind dependency* incorporates *weak-mind dependency*, since presumably the sorts of training we are capable of undergoing is dictated by our human-ness.

A third sort of mind-dependency can be distinguished. Notice that in both the case of weak or

\(^{230}\) Dennett, 1991, pg. 36
strong mind-dependency, there is a notion that the pattern being recognized is discernable-in-principle; that is, the pattern is there to be perceived, one just needs the proper equipment to pick it out. The ultra-violet patterns to which bees are attuned, for example, are there for us to ‘figure out’, given some technological or conceptual extension of our sense organs. Some mind-dependent patterns may not be discernable-in-principle in this way in the sense that they may be entirely mind-dependent or are patterns that are, strictly speaking, generated by the mind. Fictional characters, arbitrary organizing principles, some (the intuitionists like Brouwer) would say numbers and mathematical objects are all in this category. This sort of mind-dependency does not particularly concern me, but I do note the possibility.

Here then, are the three levels associated with mind-dependency:

Level I: Mind-Independent Patterns: No dependency on minds whatsoever means that the patterns at this level could be identified by some cold, objective language like that of mathematics and Ontology I. Level I (Mind-Dependency) would correspond quite directly to Level I (Algorithmic) for this reason, but not necessarily to Level II (Algorithmic) as the function $f(x)$ itself may be mind-dependent.

Level II: Weakly Mind-Dependent Patterns: patterns all humans see, but that we all see as a result of being human (statistically minor exceptions, such as autism, aside).

Level III: Strongly Mind-Dependent Patterns: patterns that all humans could see, but only some do and only with effort and the relevant training (again, statistically minor exceptions to our shared abilities aside).

The examples I have outlined would be distributed along this axis as follows:

Mathematical Objects: Level I

Natural Kinds (maple trees and tulips): Level II, perhaps Level III if training is required (which is surely the case for many species).

Artifacts: Level II or Level III, depending on the artifact. Some artifacts may only be recognized as the tools they are with sufficient training – can a layman see an electron microscope?

Complications of colour-blindness, possible spectrum-inversions, etc. are to be ignored.
**Gaming Kinds:** Level III, one needs training to see these patterns.

**Intentional Kinds:** Level II for the main part; one needs training in the sense that one needs to be in the presence of other people during childhood, but one does not need specific training - the learning appears natural or inevitable in a way that is different from a ‘weak king side’. I take it that natural language is of this type, if not the complex vocabulary associated with some specific area of expertise. Some intentional language, however, would be Level III as it does not seem to be part of our automatically shared toolkit (can all humans identify ‘intuitive understanding’ or ‘ponderously inundated by melancholy and existential angst’?).

Note here that some degrees of complexity (Mind-Dependency) are inverse to the degree of complexity (Algorithmic). Intentional patterns are all Level III (Algorithmic) but some are Level II (Mind-Dependency), and gaming kinds are the reverse. I take it that this can be taken as evidence of some sort of pre-disposition to see patterns of human behaviour in intentional terms, such as a theory of mind of some sort.

### 2.5.3 Degree of Context-Dependency

Here, I develop the point made earlier by Haugeland, regarding the degree of integration between the elements of a pattern and the larger pattern, or context, in which it is embedded. I want to define a distinction between strong and weak context-dependence.

Strong context-dependence is meant to capture two distinct (but related) ideas; in each case a kind of recursion renders the notion of independently defined individual elements as nonsensical. The first sort of recursion is the idea that the larger context/pattern is defined (in whole or in part) by the elements/sub-patterns that make up that larger context, and the elements/sub-patterns are themselves defined (in whole or in part) by the larger context. This is precisely Haugeland’s point made earlier. Thus, the elements and the larger context mutually define or re-enforce each other; that is, there is a recursive or two-way direction of dependence. The second sort of recursion is not necessarily a mutual recursion, but one that requires ever-higher levels of context for disambiguation.

Examples of the first sort are invented games such as chess, which displays strong context-dependence by definition or for reasons of analyticity; recalling Haugeland - chess pieces are
defined relative to the rules of the game, and the rules of the game are defined relative to the allowed
moves of the pieces. An example of the second is natural language, in which the meaning of a word
is not just dependent on the surrounding text (and vice versa), but each of these is also dependent
on the larger context of a shared understanding (or worldview) existing between the two people
involved in the exchange. The earlier reference to Dreyfus’ “stay close to me” is a good example.
To disambiguate the meaning of ‘close’ one needs to bring in ever-higher contextual information;
the other words, the shared environment of the speakers, the shared histories of the speakers, the
shared life or world-view of the speakers, etc. This is not the case only for synonyms: see my
“home is where the heart” example, Section 3.2.2.1. In the chess example, the regress is virtuous,
and in the natural language case it seems a more vicious regress, and one that is broken only by
what Wittgenstein calls our ‘forms of life’, and what I am calling our worldview. I shall have more
to say about this Wittgensteinian notion in the next chapter.

Weak context-dependence is meant to capture a less recursive notion of dependence, in which the
larger context can be defined independent of the elements in question, but that context does
contribute to the identification of the patterns which are elements of that context.

Here then, are the three levels associated with context-dependence:

**Level I: Context-Independent**: The pattern is identified independently of the larger context in
which it is embedded, and the elements of the pattern may also be so independently defined.

**Level II: Weak Context-Dependent**: There is some dependency of the elements on the context,
but not of such a type to generate an infinite regress or mutual reference.

**Level III: Strong Context-Dependent**: Enough dependency to generate an infinite regress or
mutual dependence.

The examples I have outlined would be distributed along this axis as follows:

**Mathematical Objects**: Level I

**Natural Kinds** (maple trees and tulips): Level II, perhaps Level I. Level II if the context is taken to
include the presence of a recognizer. This is a complicated issue, and one that is dependent on how
the natural kind falls on some of the other axes; if tulips are not Level I (Algorithmic), but Level II
or even III (Algorithmic), for example, do they exist as *tulips* without a creature to identify them as such?

**Artifacts:** Level II or III. Tools may exist as functions within larger tools, a chip may act as a memory bank only by virtue of participating in a computer. This is analogous to gaming kinds in the sense that the functions of the whole and its parts, and their relation to each other, is in a very strong sense, *designed*, defined or decided upon by the builder of the artifact.

**Gaming Kinds:** Level III, paradigmatic case of analytically defined mutuality.

**Intentional Kinds:** Level III, following Dennett, it takes another creature to adopt the Intentional Stance, and following Dreyfus on natural language, it is only by virtue of a shared world-view that we are capable of understanding each other at all.

### 2.5.4 Comments on Pattern Complexity

In the last section, I loosely tied complexity with realism, and the reason to be so loose is to point out that complexity is not a simple issue (sic). There are - as I have shown - a number of axes by which to measure complexity and they can be related in various ways.

In discussing natural kinds, Goldman says “Presumably something counts as a natural kind only if it has a prior essence, nature, or character independent of anybody’s thought or conception of it. It is questionable, however, whether such analysanda as knowledge, justification, and justice have essences or natures independent of our conception of them.” Here, Goldman is making an (implicit) link between two axes. A natural kind exists as mind-independent entity only if there is a ‘prior essence’; prior essence thus needs to be defined. This a controversial issue, but I presume would be settled by the existence of genetic identities that can be identified at Level I (Algorithmic). So, it appears that the degree of amenability of algorithmic description (AAD) can be an indicator of the degree of mind dependency, at least as far as AAD is used to identify the pattern in the cold, objective terms of independently-defined elements and computations over those

---

232 Goldman, Pathways, pg. 82
233 I do not discount other possibilities (like a Platonic Form, floating ‘out there’ somewhere) but as a physicalist committed to the metaphysical priority of Ontology I, this would be the only sort of argument by which I would be convinced.
elements. I am not arguing that AAD is a necessary condition for mind-independence, but a sufficient one.

The axis of mind-dependency is, of course, related to the pursuit of science. Science has long tried to establish as high a degree of mind-independency as possible, and does this by establishing the double-blind experiment for medicine, by establishing a strong tradition of peer-review, and by the most commonly identified method of establishing objectivity, that of the repeatable experiment; science requires that all experiments need to be repeatable, anywhere by anyone. Mind-independence is here being established as much by methodological considerations as by way of claims related directly to AAD. Feyerabend (1975, 1999) and Kuhn (1970) have each made strong criticisms of the assumptions of objectivity and mind-independence in science, and these social commentaries aside, science does still struggle with ways of keeping minds out of it; certainly the individual minds of Level III (Mind-Dependence), and ideally also the collective mind of Level II (Mind-Dependency). Hacking (1983) has spent a great deal of time defending science as objective, yet mind-dependent.

An example of an attempt to eliminate the ‘intrusion’ of mind into the empirical domain of science is given below in an excerpt from a recent New Scientist discussing the status of various large-scale patterns believed to be evidence against the current standard cosmological model of uniform background radiation;

“Some believe [the pattern] is just a figment of overactive imaginations. But evidence is growing that the so-called ‘axis …’ – a pattern apparently imprinted on the radiation left behind by the big bang – may be real, posing a threat to standard cosmology …. According to the standard model, the universe is uniform everywhere … Kate land and Joào Magueijo of Imperial College London noticed a curious pattern in the map of the cosmic microwave background (CMB) … It seemed to show that some hot and cold spots in the CMB are not distributed randomly, as expected, but aligned along what Magueijo dubbed the axis …. Some astronomers have suggested straightforward explanations for the axis, such as problems with the … instruments or distortions caused by a nearby supercluster … Others doubt the patterns very existence. … “there is a danger that once people know about the axis of evil, they start seeing evil in all sorts of sets of data” … Two independent studies seem to confirm that it does exist … analyzed the polarisation of light from 355 quasars and found that as the quasars get near the axis, the polarisation becomes
more ordered than expected. Taken together, the polarisation angles from the 
quasars seem to corkscrew around the axis. [Carlo Contaldi of Imperial College 
London]234

Here, the struggle is to ensure that the pattern is Level I (Algorithmic), as well as Level I (Mind Dependence), and all sorts of other patterns are found to justify the existence (or non-existence) of the first pattern. All these patterns, of course, become mere dots on a page that can be analyzed as a visual pattern only after an enormous amount of highly theory-laden data is analyzed; can that highly theory-laden data ever be Level I (Algorithmic or Mind-Dependence), given the collective mind-dependence of our theoretical constructions? I leave this question open.

I have advocated that a pattern’s (ontological) status as being discernable-in-principle does not hang on claims of that pattern’s AAD, although I would certainly want to say that the degree of AAD indicates a method of providing the discernability condition itself; patterns that are AAD are certainly discernable-in-principle, using just the restricted language of what I have called $x$ and $f(x)$ (the elements and computations over those elements). Hence, a degree of ADD is sufficient to establish the discernability-in-principle claim. Patterns that are not AAD (because of a high degree of mind-dependency, for example, like intentional patterns of pathos on the television screen), however, I would still claim are discernable-in-principle, and I have argued that the discernability criterion be extended to include the presence of an observer, with the attendant perspective, interests and perceptive abilities. That observer is bringing to bear the operational notion of recognition. Hence, I want to untangle the notion of discernability-in-principle (of a pattern) from the necessity of including an operational notion of recognition (in that patterns identify conditions).

So, aside from Level I patterns— which appear common across all axes as being limited to mathematical objects and the like – patterns have various degrees of complexity, along various axes, and which are variously related. I hope to have given some intuitive understanding of the number of possible ways to go about establishing what a pattern is and how its degree of complexity may be established. However, nowhere does it seem very plausible to float free of an operational definition of patterns.

I turn now to developing that operational definition.

---

234 New Scientist magazine, April, 2007
2.6 Final Definition of Pattern

Since patterns cannot be defined independently of acts of recognition (that is, there must remain an operational element in their definition, with the exception of perhaps Level I) I propose the following initial definition of a pattern, given the metaphysics of this chapter.

(A) Pattern: A pattern is whatever can be reliably picked out as a pattern, by a pattern-recognizing creature. A pattern-recognizing creature is a sentient being capable of pragmatically negotiating their way around in the universe, and the patterns are – relative to their interests – a compression of Ontology I.

The potential vacuity, circularity or even idealism introduced by the self-regulating flavour of this definition (a pattern is what is taken as a pattern) - a madman perceiving patterns everywhere, or the roulette-player constantly ‘seeing’ patterns in the run of numbers - is avoided by the introduction of the ‘pragmatically negotiating’ term; only patterns that are really there will reward the recognizer with the pragmatic reward of successful negotiation. The roulette-player will eventually lose, the madman will not be understood by his peers, etc. I don’t want to have the ‘pragmatic’ here be interpreted as having some sort of actual pay-off for the recognizer; pragmatic here really just means there is some external, empirical constraint provided to gauge the correctness of recognition.

The potential vacuity of ‘pattern’ being mere differential response, however, remains; a flower detects the pattern of the sun moving across the sky. I do not hesitate to call this a pattern - however, the patterns we are interested in contain a sense of the normative, and so I embed that pattern-recognizer in a linguistic community. That community ensures a degree of normativity in acknowledging (or questioning) speech acts related to the identification of patterns. I will have more so say about the role of language in Chapter 3 – Recognition.

I thus strengthen the definition to be the following – noting that I am differentiating now the sorts of pattern-recognition that may be performed by linguistic creatures, language bringing a normative element to the act of recognition so that it is more than mere differential response:

(B) Pattern: A pattern is whatever can be reliably picked out as a pattern, by a pattern-recognizing creature. A pattern-recognizing creature is a sentient being, embedded in a linguistic community, capable of pragmatically negotiating their way around in the universe, and the patterns are – relative
to their interests – a compression of Ontology I and has associated with it some (newly learned or culturally inherited) linguistic item(s) agreed upon as a matter of consensus by the community.

Note that I do not contend that patterns that cannot be recognized do not exist. Recognition may be at the heart of my definition, but the capacities of discrimination I take to be open-ended in possibility. The existence of systems capable of recognizing patterns of *whatever* degree of subtlety or complexity is an entirely empirical matter, but the *possibilities* of recognitional capacities are not limited by those empirical facts.
Chapter Three

3. PATTERN EPISTEMOLOGY OR ‘RECOGNITION’

This chapter fills in the irreducibly operative part of the definition developed previously; ‘recognition’ remains a part of the definition of ‘pattern’ developed thus far, and here I analyze that operational notion. There yet remains an operational flavour, however, in that what makes something a pattern-recognition system is the behaviour or abilities it exhibits; that operational flavour will be characterized, however, in detailed functional terms. Since I am here concerned with human cognition, I shall follow (and criticize) Dreyfus’ development of the pattern recognition abilities that we humans exhibit in solving the problems that we do - in direct contrast to the capabilities thus far exhibited by traditional computing systems within the field of Artificial Intelligence (AI). One field of particular interest within AI is ‘cognitive simulation’ (CS), which is an effort to mimic human intelligence or behaviour in a computing device, and a central problem of over-riding importance within CS is the (in)ability of machines to interpret natural language (note that I have previously explicitly equated natural language interpretation with pattern-recognition).

The four specific – and somewhat inter-related - functional abilities here associated with pattern-recognition Dreyfus calls fringe-consciousness, ambiguity tolerance, essential/inessential discrimination and perspicuous grouping. Part and parcel of the arguments made in this chapter is to claim that these abilities are responsible for the more general cognitive tasks of natural language interpretation and concept identification. Taking the presence of these abilities as a group in some system, I shall call adopting a pattern-rec stance toward that system.

In the previous chapter I argued for a strong metaphysical conclusion; patterns retain an irreducibly operational flavour because of the nature of ontology, because the nature of the world itself is such that it is not composed of discrete facts waiting to be agglomerated into discrete patterns. Dreyfus calls the assumption that it is so composed, the ontological assumption. My distinction between Level II and III (Algorithmic and Context-Dependence) patterns is really meant to be the high-water mark, delineating the validity of that assumption. Here, although I shall continue to argue for this same strong metaphysical stance, the burden of proof is much lower than on the metaphysical scales, since here I speak of how it is we think about the world. Our epistemology may be shaped by pragmatic limitations, even though our metaphysics is not. Even if, for example, it is not the case that the Level II/III (Algorithmic) can be maintained, it need only be true that the computational complexity of Level II is high enough that our poor brains cannot perform the requisite computing
steps. That being the case, then claims about our psychology hold independent of the validity of the stronger, metaphysical picture I have tried to justify. I urge the reader to keep this milder demand in mind, since it is ultimately an epistemically-flavoured conclusion that I seek.
3.1 The Pattern-Rec Stance and a New Computing Primitive

In adopting the *pattern-rec stance* I make no assumption during these initial discussions about the class of physical systems that count as pattern-recognition systems; all of the cognitive abilities and behaviour associated with this account shall be couched in functional terms. The *pattern-rec stance* then, is meant to capture a class of systems in functional terms. I argue elsewhere, of course, that one type of physical system that *cannot* account for these functional abilities, due to the nature of the patterns themselves (Chapter 2) and due to the failure of the Language of Thought thesis (Chapter 4), are symbol-crunching Turing machines. A positive argument for one class of system that *can* account for these abilities (C-Nets) can be found in Chapter 5.

This set of functional abilities - that, I shall argue, are displayed by any system capable of interpreting natural language and identifying the concepts of interest - is what allows one to adopt what I am calling the *pattern-rec stance*; *We exhibit particular functional properties – call them pattern-recognitional abilities - that are necessary (but not sufficient) for a system to be capable of recognizing the sorts of patterns that we do.*

Note that I am making no claim that C-Nets are the *only* sort of system that could account for pattern-recognition behaviour. No account in physical terms, in my view, captures the full relevant class of all pattern recognition systems except *per accidens*; i.e., the connectionist story is demonstrably compatible with the *pattern-rec stance* toward human cognition, but there may be any number of other potential implementation stories. That said, the *pattern-rec stance* does rule out some implementation stories, e.g. symbolic computing.

What I claim about the *pattern-rec stance*, is that in adopting such a stance toward a system one can take for granted the existence of a certain sort of ability as a computing primitive. Just as a Turing machine takes the symbolic–syntactic link, and the physical laws under which those syntactic items operate, as primitives of the system (see Section 4.2.3 for details), so too the pattern-rec system has its primitive. That primitive I call *similarity-for-free*; by this I mean that regarded from a suitably functional perspective, *similarity between objects* can be regarded as a basic, primitive and irreducible sort of computing function - the prior existence of identity conditions is not required to judge similarity. C-Nets are held as an example of such systems, but arguments to that effect are not provided until Chapter 5.
Dreyfus himself is optimistic about the connectionist approach itself, and the associationist flavour he likened in earlier writing to a hologram; “it is important to realize that there are physical systems that can detect similarities without using any features and rules at all. ... Thus holograms can act as virtually instantaneous similarity recognizers ... In recognizing similarities that way, the question ‘similar with respect to what?’ does not arise.”\textsuperscript{235} Here, I take it that since ‘with respect to what?’ is moot, similarity is being suitably (to me) primitive. By the time the later edition of What Computers Still Can’t Do was published in 1999, that holographic analogy had been cashed out “by the rapidly growing ranks of neural-network modelers.”\textsuperscript{236} In Dreyfus & Dreyfus, the hologram is a useful analogy by which to understand connectionist networks, and the point being made is “the information processing computer is not the only physical system that can exhibit mindlike properties and that other devices, such as holographic patterns recognizers, may be closer analogies to the way the mind actually works.”\textsuperscript{237} Just so, and it is Clark whom I shall follow in bringing out this new computing primitive that more accurately reflects the mind than those of the Turing machine. I note here that the over-riding functional characteristic I am calling \textit{similarity for free} is meant- at this point - to be suitably abstracted from any particular implementation details.

\textsuperscript{235} Dreyfus & Dreyfus, 1986, pg. 87
\textsuperscript{236} Dreyfus, 1999, pg. xiii
\textsuperscript{237} Dreyfus, 1999, pg. 61
3.2 (Human) Pattern-Recognition Traits – or – Why We Are Not Computing

One of Dreyfus’ primary concerns is to demonstrate how human cognitive abilities are not currently, nor can they be, implemented on digital computer. For Dreyfus this is a problem he calls “Cognitive Simulation”, and denying that it is possible to simulate the sorts of cognitive skills that are an everyday occurrence for we humans is a part of a larger criticism of Artificial Intelligence (AI) in general, which he argues faces severe performance constraints that are a direct result of the symbolic computing backbone that is the assumed architecture of that larger project. As noted earlier, pattern-recognition is the fundamental feature that underpins most interesting human cognitive behaviour that the AI community attempts to simulate; “the resolution of the difficulties which have arrested development in game playing, language translation, and problem solving presupposes success in the field of pattern recognition” [emphasis mine].”

Showing that AI has not lived up - as a matter of empirical fact - to the claims and hopes of its many pioneers (like Minsky, Simon and Newell) is a long, detailed argument that is the focus of Dreyfus’ work, and I shall not concern myself with that argument here but shall accept it as largely true. Instead, I shall focus on the reasons Dreyfus gives as to why AI systems “lack[ed] the common sense of a four-year old, and [why] no one knew how to give them the background knowledge necessary for understanding even the simplest stories.” Four distinct cognitive abilities underpin our ability to perform the sorts of recognition that we do, and each of these abilities cannot, in principle, be instantiated in a symbolic computational device; the use of fringe consciousness, ambiguity tolerance, essential/inessential discrimination and perspicuous grouping.

A point that is key to Dreyfus’ line of argument is that each of these cognitive skills is what is actually being exercised in human cognitive activity, as opposed to some other competing (and computable) sort of cognition that is assumed to be taking place. In other words, Dreyfus is making a case for how it is that we actually cognize, and each particular skill is held up as an alternative to what Dreyfus accuses his AI detractors of assuming is taking place. Much as I shall accuse Fodor in Chapter 4 of making unwarranted assumptions about how the human psychological space is best described, Dreyfus is accusing his detractors of holding mistaken

238 As quoted earlier, Dreyfus, pg. 97
239 It is telling to note that many successes in modern pattern-recognition systems are due to a shift in computing platforms from a symbolic system to one that models connectionist networks – see Chapter 5 for details.
240 Dreyfus, 1999, pg. 108
assumptions regarding our psychology, and in each case he is offering a positive account of a viable alternative.

Thus, each type of cognition is meant to provide an alternative account to the computable function that is assumed by the AI folk, and each one is brought out by reference to an example field of Cognitive Simulation; *fringe consciousness* is meant to replace many heuristically guided searches and is applied to gaming, *ambiguity tolerance* is meant to replace a kind of context-free precision and is applied to language translation, and *essential/inessential discrimination* is meant to replace a comprehensive search by trial and error and is applied to general problem solving (GPS). Although Dreyfus focuses on a single field of cognitive simulation to make his point, it will be my contention that each of these cognitive skills applies across a broad spectrum of cognition, particularly right across much of the *epistemic continuum*. Note that these are not mutually exclusive categories, but are a guide to types of cognition; fringe consciousness, in particular, pervades all three.

*Perspicuous grouping* is a kind of culminating or over-riding description of the sort of cognition of interest, an ability to perform classifications over the types of patterns that we do in fact recognize, and is meant to replace classification of events or objects by a list of character traits. This distinction – the *method* by which classification takes place – is a key differentiator of the alternative view (to cognition as computation) that I am developing. Note that it is not classification *per se* that is being denied here, rather *classification by means of a list of character traits*. Classification here becomes an ability to perform an act in a particular context – functionally defined in empirical terms - rather than being the specified act of computing over some pre-defined list of traits or elements (this point was also emphasized in Chapter 2). The analytic project of providing necessary and sufficient conditions for concepts – and the analogous internal machinery associated with Epistemic Romanticism and Intuition Driven Romanticism (see Section 1.2.2) - is an example of the sort of classification here being denied. Recall that family resemblances were the most complex sorts of patterns; membership is open-ended in a deep and interesting way, and the members may not have any set of features common to all. These are the patterns of real interest here, and I had pushed the idea in the previous chapter that intentional concepts fall under this sort of category. Dreyfus’ point is that all three of these sorts of cognitive skills are required for the *perspicuous grouping* associated with these patterns.
3.2.1 Fringe Consciousness

Dreyfus inherits from William James the term “fringes of consciousness”, and – following Michael Polanyi – identifies fringe consciousness with items inhabit the margins of our consciousness and of which we are not overtly aware, but which never-the-less affect how our perceptual fields are conceptualized. In other words, fringe consciousness is a kind of organizing force that contributes directly to the sorts of patterns that we see; “This power resides in the area which tends to function as a background … it extends indeterminately around the central object of our attention … Seen from the corner of our eyes, or remembered at the back of our mind, this area compellingly affects the way we see the object on which we are focusing”\textsuperscript{241}. Fringe consciousness involves the processing of information “which is not explicitly considered or excluded”\textsuperscript{242}.

For Polanyi, the claim implies that conceptual knowledge affects both the quality and type of perceptual experience, and much ado has been made - by contemporary critics of the traditional view of a perceptual field as a sort of processed data grid - of the idea that our background knowledge is a literal part of the direct perceptual experience itself (see Noe, ed., 2002, for an extended discussion). The awareness of the solidity of a house means, for such critics, that we do not see a façade but see, rather, a solid three-dimensional object. Extending this notion, one may legitimately ask the question: does a trained historian see ‘merely’ a church-shaped building, or do they see a socio-historical object, imbued with religious passions, political intrigue, and the like? Much hinges on the meaning of ‘see’ here, and an analysis of fringe consciousness is meant to flesh a part of that meaning out. By itself, however, the idea that items outside our direct awareness affect our perception is not controversial, nor does it contribute any interesting degree of complexity to the cognition taking place during pattern perception – what is important is the relationship that holds between information that exists on the ‘fringe’ and the patterns being recognized.

That our perceptual field is ‘chunked’ into objects and events that are interesting to us is a familiar phenomenon, and is largely the basis of gestalt psychology, but here the interest is in how the surrounding perceptual data is used to organize the experience. The claim Dreyfus is making is that fringe consciousness is being used in a way that is distinct from the familiar heuristics (or analytics) by which we make further (conscious) use of the organized data. That is, the perceptual data at the fringe influences the perceptual content by non-heuristic means, although the conceptual

\textsuperscript{241} Polanyi, “Experience and Perception of Pattern”, 1968, pg. 214.
\textsuperscript{242} Dreyfus, 1999, pg. 107
data resulting from the perceptive act may itself be used in a heuristic (or analytic) manner. Dreyfus’ argument “while not a proof that unconscious processing is nonheuristic – does put the burden of proof on those who claim that it is or must be.”

The best developed example is one of chess. Many games, such as tic-tac-toe or checkers, can be quite easily programmed (a fact to which I can attest) and the reason why it is so simple to program these games relates to that which separates Level I and Level II patterns (Algorithmic – See section 2.5.1). Chess - “although decidable in principle by counting out all possible moves and responses” - soon hits a kind of combinatorial explosion, and the practical solution in chess programs is to find suitable heuristics to limit the exploding path of choices that face a simple ‘counting out’ (checking all possible future moves on each side to make a present decision). The question of relevance here is: is this what humans are doing? Do humans use identifiable heuristics in chess-playing?

I take it that the force of Dreyfus’ argument is amplified by choosing chess as the target. The question is of particular importance in chess because, complicated as it might be as a game, it is just that - the simplified environment of an artificial game - and one that is understood (and studied) well enough to be able to provide an answer to our question. The kind of complexity associated with the situations we are capable of dealing with – our world is of people, natural kinds and things like a ‘good pick and roll’ – is much more open-ended than the confines of the artificial world of chess. If it can be argued that we do not use heuristics for chess, so much the worse for the fate of heuristics in our everyday cognitive skills.

It is well-established that chess-masters chunk the board differently than a novice, a fact established by a number of perceptual tests. After a brief exposure to a chess-board set up to a location within a real game, chess-masters are “able to reproduce … board positions perfectly, and performance degraded appreciably with decrease in chess ability … perceptual abilities and organization were an important factor in very good play”, but if the board is set to a random mess, the chess-masters lose that ability. I take it that the fact that chess-masters chunk the board better (and differently) than novices is uncontroversial. The question remains, however: why does this mean the chess-master is not merely using different and more advanced heuristics?

243 Dreyfus, 1999, pg. 106
244 Dreyfus, 1999, pg. 101
245 Simon & Newell, quoted in Dreyfus, 1999, pg. 105
As I see it, there are actually three separate arguments made at this point, and each can be made independently. Success of these arguments provides additional means by which Level II and III (Algorithmic and Context-Dependent) patterns can be distinguished. Dreyfus makes these arguments through the example of chess, but I emphasize that there is no reason not to think that fringe consciousness does not permeate the entire *epistemic continuum*.

### 3.2.1.1 Accumulated Past and Present Actions as Context

It appears that the perceptual chunking of the board is dependent on two factors; one being the usual “familiarity with the overall chess pattern”\(^{246}\) and the other being “the past moves of this particular game”\(^{247}\). Some of the patterns that are being perceived (weak king side, “lines of force, the loci of strength and weakness”\(^{248}\) are dependent on the playing patterns of that particular opponent on that particular day, as ‘weakness’ or ‘strength’ is obviously relative to a player. One person’s weak king side may be another’s trap.

Dreyfus draws the lesson that it is both the past experience of the games played and current game being played that is being accumulated and put into cognitive use. This ‘accumulated past and present actions’ is something that humans take into consideration automatically, whereas there is “no chess program which even tries to use the past experience of a particular game in this way … [due to combinatorial explosion] … what is needed is a program which *selectively* carries over from the past just those features which were significant in the light of its present strategy …”\(^{249}\) Generalizing this idea, it is the “background of past experience … [that determines] what shows up as a *figure* and captures a player’s attention”\(^{250}\). So here, we have the problem of a (very, but not infinitely) open-ended possible relevance of all past moves to figure in the shaping of the current situation and “this gestaltist notion of figure and ground has no place in explicit step-by-step computation”\(^{251}\).

\(^{246}\) Dreyfus, 1999, pg. 105

\(^{247}\) Dreyfus, 1999, pg. 105

\(^{248}\) Dreyfus, 1999, pg. 105

\(^{249}\) Dreyfus, 1999, pg. 105

\(^{250}\) Dreyfus, 1999, pp. 105-6

\(^{251}\) Dreyfus, 1999, pg. 106
At this point I posit that a better example to use might be poker\textsuperscript{252}. The chess example is fraught with potential computational explosion, and it can always be argued that what is at issue is merely a Level I/II (Algorithmic) distinction, a distinction of mere in practice possibility – which, while highly relevant – is not as strong as the Level II/III distinction, a distinction of possibility in principle. See the next Section (3.2.1.2) for a standard response to this criticism. Poker has a much simpler computational structure, and so has no pragmatic problem of ‘counting out’, so the problems faced in duplicating human performance is much different.

Any good poker player can memorize the rough probabilities of getting various hands, and can bet accordingly. The problem in poker is bluffing. Any player can bluff anytime, and most bets are won before the hands are actually shown. The point here is that bluffing is a part of the context, and needs be taken into account. In real life, both past and current betting patterns as well as body language need to be taken into account. In on-line versions, only past and current betting patterns need be taken into account. There are human experts (it is not luck, although that is a factor) who consistently win both on-line and in real life, and there is – as far as I know – no computer that can consistently win\textsuperscript{253}.

The betting patterns by themselves are sufficiently complex that there are (even without the added complexity of human interaction) - as a matter of empirical fact - no computers that can pick them out, but there are people who can pick them out. Yet, the degree of computational complexity is tiny, so it seems not to be an issue of combinatorial explosion. It appears to be a cognitive ability of a different sort, although I freely admit it may not qualify as fringe consciousness, however, depending on the phenomenal availability of the ‘figuring out’. The distinction would be – by definition – whether the “information, rather than being explicitly considered remains on the fringes of consciousness and is implicitly taken into account”\textsuperscript{254}, and I leave that question open for now.

Note that what the poker example provides is the merest hint of the intentional (in the form of bluffing in the interpretation of that bluff), and note also that any advantage of computational

\textsuperscript{252} I presume the reader is familiar with the game, but basically: best 5 cards wins, from a hierarchy of hands. Typically players hold two cards, and five get laid out on the table (which all players share). Players bet at various times, before all 5 cards are down, and so present strength and weakness by the size of the bets. They may be lying (bluffing), or telling the truth. Normally, someone bets enough that other players back down (fold), and only occasionally does it come to a ‘showdown’, where cards are laid down and best hand wins. The point here is that bluffing is a big part of the game.

\textsuperscript{253} If there were, the on-line world of poker would be shambles, as someone would simply play with that program.

\textsuperscript{254} Dreyfus, 1999, pg. 106
prowess is immediately brought to naught (if I am indeed right about humans being better poker
players, admittedly). I take it that the Level III (Algorithmic) patterns in which we have more
substantial interest involve *fringe consciousness* in a more uncontroversial way. Attributing mental
states, beliefs, hopes and wishes to other people who act within a great wash of human activity
happening all around them, the overriding context of which is to make sense of that behaviour in a
framework of rationality, surely must be a case where “complexity is such that global awareness
would be necessary to avoid an overwhelming exponential growth of possibilities to be counted” and
a good reason to think that such information is, indeed, “implicitly taken into account”.

3.2.1.2 *Heuristics Top to Bottom?*

A reasonable response to the above (chess-based) argument could be to argue: “*All that mysterious
underground non-heuristic stuff is just an assumption, why can’t I just say it’s counting out and
heuristics right the way through – from the unconscious bottom to the conscious top - and the
unconscious stuff is just really fast computing that one is simply unaware of? Why must I accept
that something different is going on?*” It may be that the human player under consideration is
“not aware of having explicitly considered or explicitly excluded from consideration” any of the
parts of the board, or possible moves, but that does not mean she has not done so, albeit
unconsciously.

Well, firstly, I would caution a person relying on this argument against treating the computational
combinatorial explosion so lightly. The problem going from Level I to Level II (Algorithmic) is not
trivial, and the number of neurons we have with which to compute may be large from one
perspective (as an absolute number) but not from another (compared to computing steps). Although
I keep pushing for the theoretical distinction associated with an *in-principle* impossibility of
algorithmic description, the very real pragmatic concerns associated with the *in-practice*
impossibility also places severe restraints on our *psychology* if not our *metaphysics*. Aside from
that cautionary note, however, the *empirical* evidence is actually in favour of the ‘something
different’ thesis.

The form of the counter-argument can be put as a question: if it were heuristics all the way down,
why is it not heuristics all the way to the top? Even grand-masters need to contemplate the next best

---

255 Dreyfus, 1999, pg. 107
256 Dreyfus, 1999, pg. 106
257 Dreyfus, 1999, pg. 103
move, after they have ‘zeroed in’ on the relevant and finite set of moves to consider. The question becomes, “Why, if the unconscious counting is rapid and accurate, does he resort to a cumbersome method of slowly, awkwardly, and consciously counting things out at [the point at which he has recognized some advantageous or interesting position]?” There does indeed appear to be two distinct types of cognition, the ‘zeroing in’ on what’s important, and the deliberative process from that point forward. Also, the more intentional poker example does not fall easily to this argument; if the betting patterns that correspond to the possibility of bluffing were amenable to a computational description, I do not think that there would be human domination of the very lucrative on-line poker world. Neither of these counter-examples count as a proof, admittedly, but do “put the burden of proof on those who claim that [heuristics are involved]”

It must be admitted, however, that such arguments can be quite easily constructed. One may argue that it may well be the case that fast under-ground heuristics (rather than counting out) might be necessary to zero in on the relevant portion of the board, and relativize various patterns to the current context and player (as ‘weak’, etc.), but those heuristics are fast and general and by their nature can only in very broad strokes provide a next move. Those heuristics break the pattern up into manageable chunks, but considering a strategy as to moving an individual piece, that is a different sort of task that has not been taken in (or trained up) by the underground processes and must therefore be counted out. This seems a reasonable enough response (although I do not think it sufficient given the arguments related to the metaphysics of pattern in the previous chapter), and I turn now to another counter-argument - given as the Ambiguity of sub-Pattern Identity argument in the next section.

3.2.1.3 Ambiguity of sub-Pattern Identity

In the spirit of Haugeland’s inter-contextual dependency of patterns, I take it that the chunking that is taking place in fringe consciousness does not have the (computational) luxury of dealing with sub-patterns that are independently defined. Just as a ‘weak king side’ has the identity it does by virtue of participating within a larger context that includes the abilities demonstrated by the other player – that’s why it’s weak – and just as a rook is what it is by virtue of participating within the larger context that is the game of chess – that’s why it’s not just a castle-shaped thingy – so too, whatever sub-patterns that allow for the gestalt-like chunking of the board play the role they do in

---

258 Dreyfus, 1999, pg. 106
259 Dreyfus, 1999, pg. 106
the chunking process by virtue of their *functional significance*, which depends in turn on a surrounding context. I do not know my chess well enough to develop this argument, but it seems if Haugeland’s *strong context dependence* has any force, it surely adds another degree of computational complexity to any heuristics that might be used. The problem is that whatever are assumed to be treated as independent elements in the heuristic approach, whatever computational elements are to be thought of as primitive in the generation of *gestalt*-like patterns, are not defined independently of the functional pattern in which they are to participate. Positing heuristics in fringe consciousness does not cut that particular Gordian knot.

This is not a proof, of course, but merely a warning that any assumption that *fringe consciousness* is composed exclusively of unconscious heuristics and counting out must deal with a suitably complex, inter-dependent set of entities that display the self-referential nature of *strong context dependence*. This open-ended ambiguity of the individual elements of a pattern provides a segue-way to the next type of cognitive skill, in which ambiguity plays a central role.

### 3.2.2 Ambiguity Tolerance

*Ambiguity tolerance* is brought to the fore in the analysis and translation of natural language, and so is of more direct interest here since the type of philosophical analysis under discussion is one in which a precise definition of knowledge (or some other concept) is being tested against textual descriptions of various situations. Attempts to build algorithmic language translators have highlighted the form of the problem; one needs to generate an “ability to deal with situations that are ambiguous without having to transform them by substituting a precise description.”

It is not the case, Dreyfus argues, that either the “order of the words in a sentence … nor … the surrounding words – the written context” provide for an adequately precise semantic appraisal. Cognitive activity is taking place within *fringe consciousness* since “the user of a natural language is not aware of the many cues to which he responds in determining the intended syntax and meaning” and the thrust of Dreyfus’ argument is to show that a computational or heuristic appraisal of such cues cannot be taking place – that is, a computational or heuristic approach is not being activated within fringe consciousness since such an approach cannot provide the necessary disambiguation.

---

260 Dreyfus, 1999, pg. 107
261 Dreyfus, 1999, pg. 107
262 Dreyfus, 1999, pg. 107
Dreyfus provides **two such arguments** to this end, which are provided in the next two sections (Sec. 1 and 2); the **first** is based on the claim that there appears to be an infinity of individual facts to which one must appeal (complemented by an attempt to treat the surrounding context as a kind of fact), and the **second** establishes the coherence of there being some ‘ultimate’ context to which one makes appeal, that stops the sort of regress associated with the first argument. Dreyfus shall reject the first argument as insufficient to establish the claim, however - even though the conclusion adopts a sympathetically pessimistic view of computation as sufficient for cognition - since there is a mistaken underlying assumption about the nature of ‘fact’. Once that underlying assumption is rejected, a more holistic notion is established in the second argument, of an ultimate context to which one makes implicit appeal, and the recognitional capacities associated with that context – **situational recognition** - are developed.

### 3.2.2.1 An Infinity of Facts vs. Recognizing a Situation

There are here **two sub-arguments** that are each based on a similar underlying – and very basic and commonly adopted – assumption, which Dreyfus calls the ‘ontological assumption’; that is, “the world can be analyzed as a set of facts”\(^{263}\). As we saw in Chapter 2, the *ontological assumption* asserts that the world is composed of independently individuated facts, and this is precisely what both Dreyfus and myself are arguing is false; this is the heart of my stronger ontological claim, but remember that as a claim about our *epistemology*, I need only the weaker claim that our psychology does not take the world as such. That said, it is the stronger claim I wish to defend. Here, I relate Dreyfus’ rejection of the first sub-argument to the distinction between Level II and III (Algorithmic), and the rejection of the second sub-argument to the distinction between Level II and Level III (Context-Dependence). It is in the rejection of these sub-arguments that the recognitional capacity of a creature for *situations* in which that creature is embedded - **situational recognition** - is introduced. The nature of this recognitional capacity is then further developed in the next section.

The **first sub-argument** – attributed to Bar Hillel – is meant to establish a conclusion sympathetic to Dreyfus’ (and my) general line of argument; that computing over individual facts is not sufficient to account for the cognitive skills deployed in disambiguating textual meaning. For Bar-Hillel this is due to the potential infinity of facts to which one must refer. Dreyfus shall reject this argument as insufficient, however, on two grounds. First, it is unclear what is meant by an ‘infinite’ number

\(^{263}\) Dreyfus, 1999, pg. 218
of facts, nor is it clear that – even granting the need for some sort of an infinity of facts – that the sort of impossibility being established is one of *in-principle* rather than *in-practice*. It is the in-principle impossibility of computing various patterns that brings out the distinction between Level II and Level III (Algorithmic) patterns, for me, and it is that distinction that we are interested in sustaining (and it is that distinction that Bar-Hillel *fails* to maintain).

The second sub-argument - attributed to Katz and Fodor - attempts to limit the potential infinity of facts by establishing a ‘socio-physical’ setting as an independent disambiguating context. Here, Dreyfus rejects the argument on the basis that the context is itself being treated as a kind of independent fact, and thus reduces the problem to another instance of the (mistaken) ontological assumption. I shall show that treating the context as an independent fact, albeit one that contributes to a disambiguation of meaning, is the sort of weak context-dependence associated with Level II (Context-Dependence) patterns, and is not sufficient to account for the strong context-dependence associated with Level III (Context-Dependence) patterns.

Bar-Hillel argues that in order to suitably disambiguate the meaning of an ambiguous sentence, the number of potential facts to which one must make reference is “in a certain very pregnant sense, infinite”. For Bar-Hillel, this “amounts to an almost full-fledged demonstration of the unattainability of fully automatic high quality translation”. While it may be true that the number of facts is indeed infinite, Dreyfus points out that this is insufficient to show that the sort of ‘unattainability’ here is of merely the pragmatic sort rather than a problem in principle. Two problems lie here; what Bar-Hillel means by a ‘pregnant infinity’ is not clear, and underpinning Bar-Hillel’s claim is a simplistic notion of fact as something that exists independently from context.

Bar-Hillel’s example is to disambiguate the sentence ‘The box was in the pen’, taken from the larger text ‘Little John was looking for his toy box. Finally he found it. The box was in the pen. John was very happy’. Here, the point is that one must be able to take into account further facts – outside of the text – that are a part of our everyday understanding of the world, such as “his knowledge that the relative sizes of pens, in the sense of writing implements, toy boxes and pens, in the sense of playpens…are such that when someone writes under ordinary circumstance and in something like the given context … he almost certainly refers to a play pen and most certainly not to a writing pen”. Bar-Hillel is arguing that since our knowledge consists of an infinite number of facts, and since that degree or type of knowledge is required to disambiguate meaning, and since

---

265 Dreyfus, 1999, pg. 215

112
that number of facts cannot be programmed, a computer cannot perform the required disambiguation.

Dreyfus points out, however, that given a finite number of primitives and some computation over those primitives, a potentially infinite number of facts can indeed be generated\textsuperscript{267}. This may be true, but one must be careful to distinguish between two possible scenarios – and Dreyfus is not so careful. In the first, an infinite number of facts are actually created (which is not possible in practice, but is possible in principle) and in the second, it is from a potentially infinite set of facts that those created are a sub-set (which is possible both in principle and in practice). I take it that it is the second scenario that interests Dreyfus, and from this perspective we can ask what sort of limitation Bar-Hillel has established. There remains the problem of relevance or significance – which facts of the potential infinity are relevant, or must they all be considered (see below)? - but I shall ignore that problem for now. Assuming that such computing primitives could be found, it is the distinction between an in-practice or an in-principle computing limitation that renders the distinction between Levels II and III (Algorithmic) tenable; “... if ‘unattainable’ means [an in practice limitation] ... then the point is well made ... But if [it] means theoretically impossible – which the appeal to infinity seems to imply – then Bar-Hillel is claiming too much ... for ... from a large number of facts and rules for concatenating them, such as the laws of physics, it could produce further ones indefinitely”\textsuperscript{268} If all that is needed are facts (even an infinite number of facts), there would be no reason to deny the possibility that “only a finite number of relevant facts need be considered”\textsuperscript{269} (and that, it seems to me, would mean I must acknowledge the elimination of the distinction between Level II and Level III (Algorithmic) patterns).

I have argued that such a distinction does exist, however, and the way that I understand Dreyfus to bring it out makes use of the distinction between what I have called Level II and Level III (Context-Dependence) patterns, or between patterns that are Mildly Context-Dependent and Strongly Context Dependent. Recall the distinction was one highlighted by Haugeland, and patterns that are

\textsuperscript{266} Dreyfus, 1999, pg. 215
\textsuperscript{267} This is the thrust of the psychological attribute of productivity on which Fodor places so much emphasis (see Section 4.1.3) and which he believes can only be accounted for by a computational model of mind and a set of mental primitives that comprise a Language of Thought. The two arguments have the same flavour, but are distinct, in that productivity is the attribute of producing a potential infinity of thoughts from a finite base of primitive elements, and here it is the taking into account of a potential infinity of thoughts. It is, of course, by producing facts that they can be taken into account, but there are important differences, mainly related to the problem of relevancy. Productivity, on Fodorian terms, either assumes a degree of relevancy (the thoughts produced from the potential infinity are relevant to the task at hand) or faces the problem (how is it that only relevant thoughts are produces from the potential infinity?). Here, it will be the case that relevancy is derived by ‘situational recognition’ (see below).
\textsuperscript{268} Dreyfus, 1999, pg. 217
Strongly-Context Dependent have associated with them a kind of infinite regress in deriving the relation between sub-pattern (fact) and pattern (context). Merely bringing context into the picture (in the form of Weak Context Dependence) is not enough to maintain the Level II – III (Algorithmic) distinction, as reference to the Katz & Fodor argument will make clear.

Katz & Fodor agree that an infinity of facts must indeed be taken into account, and that it is the full spectrum of a speaker’s knowledge which is relevant in disambiguating meaning – “practically any item of information about the world is essential to some disambiguations ... [and so one cannot] distinguish between a speaker’s knowledge of his language and his knowledge of the world”\textsuperscript{270}. This is not a happy picture for them, and they attempt to resolve the ‘infinity’ problem by an appeal to context, albeit a limited one – “a limited theory of [fact] selection by sociophysical setting can be constructed.”\textsuperscript{271} So a limited number of facts can be chosen or made significant from the potential infinite number of facts that constitute our knowledge, by making reference to a constructed sociophysical setting. Thus, facts do not – for Katz & Fodor – gain significance independent of context as it is the sociophysical setting that is meant to act as the larger context (pattern) into which the facts (sub-patterns) sit and gain significance. Note, however, that there is no interesting mutual dependence between fact and context here, and certainly no possibility of an infinite regress, and so the kind of context-sensitivity is merely one of Level II (Context-Dependence). The sociophysical context is “what determines the significance of the facts”\textsuperscript{272}, but the sociophysical context is itself assumed to act, or be constructed, such as to be independently unambiguous.

Thus, the appeal to context has allowed for a ‘setting’ to play a role in the disambiguation of a fact, but a deeper problem remains, for “the setting is itself identified by features which are facts, and functions like a fact in disambiguation”\textsuperscript{273}. This is why I claim that this appeal to context is merely Level II, Mildly Context Dependent, and also the core of the reason why such an argument does not fly; the problem that remains is that the context may be here given a role in determining the significance of a fact, but the context is itself being treated as if it were an independently defined entity and not in any way determined by the fact to which it contributes a disambiguating role. This is simply not the case for many patterns of interest, as there is a “reciprocal determination”\textsuperscript{274}.

\textsuperscript{269}Dreyfus, 1999, pg. 218
\textsuperscript{270}Katz & Fodor, 1964, pg. 489
\textsuperscript{271}Katz & Fodor, 1964, pg. 490
\textsuperscript{272}Dreyfus, 1999, pg. 216
\textsuperscript{273}Dreyfus, 1999, pg. 217
\textsuperscript{274}Dreyfus, 1999, pg. 220
between fact and context; “We seem to understand the situation in terms of the meaning of the words as much as we understand the meaning in terms of the situation” \(^{275}\).

What we humans are doing is not differentiating between fact and context – that is an artificial distinction “dictated by the nature of a digital machine” – what we are doing is “recognizing situations” \(^{276}\) and it is “our sense of the situation which enables us to select from the infinity of facts the immediately relevant ones, and once those relevant facts are found, enables us to estimate their significance.” \(^{277}\) What we are doing, then, in determining the meaning of words, is recognizing the situation.

An example will help make this clear. Take the sentence ‘Home is where the heart is’. Even leaving aside the metaphoric nature of the two words ‘home’ and ‘heart’, attempts to get at the meaning of these words without simply recognizing the relevant situation in some holistic manner are hopeless. Even if ‘Home’ were merely a physical structure, we would already come up against a number of unanswerable questions: does it include the back yard? Does it have windows and is it warm there? Does a van count as a home? Does one have to own the home, or can one rent? What does ownership and rent imply, what sort of social and ethical constructs must one take into account to resolve ownership or right of residence? Home is not merely a physical structure, however, it is not merely an apartment, tent or house. It is here a place associated with belonging. Belonging has another set of open-ended questions: Am I loved there? Who loves me and what on earth is love? What gives me my sense of belonging, is it a calmness or an excitement? We haven’t even gotten to ‘heart’ yet! What on earth is that and how would I identify it by a series of independent facts and contexts?

The point of all this is to say that one cannot dodge the infinity bullet by simply referring to a context, unless that context is somehow taken to be simply recognized and not computed itself. What we humans do is recognize situations, and facts gain significance by virtue of that cognitive skill. What that cognitive skill does not consist in is computations or heuristics over independently defined facts and contexts. The relationship to previous talk of ‘forms of life’ should be obvious – it is by virtue of living our forms of life that we are capable of recognizing contexts in meaningful ways. Recognizing a situation depends on the involvement or participation of a creature in their form of life. Such a recognition is implicit in living a life, it is not explicit nor overt, but it is a kind of over-riding sense of being and of relevance and ‘situatedness’ that provides the meta-context in

\(^{275}\) Dreyfus, 1999, pg. 220  
\(^{276}\) Dreyfus, 1999, pg. 218
which other patterns are to be placed. Just as rationality is the framework by which intentionality is
to be understood, a form of life is the living framework by which other patterns are made possible.
It enables the ultimate disambiguating context – what I would want to call a ‘worldview’ – and it is
my position that it is as much a property of the observing creature and the life they lead as it is a
property of the world outside of that creature.

It must be admitted that not all words involve this degree of ambiguity, of course, but it is the higher
levels of complexity in which we are interested - particularly those patterns associated with
intentional terms - and I am claiming that these terms do, in general, exhibit this sort of open-
ended context-dependence. As stated earlier, I am not convinced by those arguments that attempt to
establish that knowledge is a natural kind of a distinctively unambiguous flavour.

3.2.2.2 Ambiguity and Particular Situational Recognition

The thrust of the preceding argument is to claim that there are an infinite number of possible cues
relevant to the task of disambiguating linguistic meaning, and this “suggests that perhaps it is not a
question of cues at all”278. The alternative view developed here is based on the idea that a more
holistic sense of the overall context is responsible for organizing our perception of language, and
that holistic sense is what limits the possible infinity of cues to some manageable subset that are
relevant to some given, particular situation. That limitation is what Dreyfus means by ‘tolerance’. I
take it that there are two notions of ambiguity tolerance that one must be careful not to confuse.
Being tolerant of ambiguity in the restricted sense of being capable of eliminating that ambiguity is
not at all the same thing as being able to tolerate the continued existence of the ambiguity itself.

For Dreyfus, the whole problem of disambiguating the sounds, or words, as independent elements
that are the “same in each context and must be disambiguated is a problem for computers, not
human beings. [emphasis mine]”279 Ambiguity is tolerated in the sense that there is no need to
assume the strong requirement that “meaning is never unambiguous in all possible situations – as
if this ideal of exactitude even makes sense [emphasis mine]”280. Meaning is disambiguated
relative to a particular context (an assumption in this section is that this cognitive task is performed
automatically within fringe consciousness).

277 Dreyfus, 1999, pg. 218
278 Dreyfus, 1999, pg. 108
279 Dreyfus, 1999, pg. 108
280 Dreyfus, 1999, pg. 108
However, I shall further argue in Section 4.4.1 that – independent of the automatic and non-computational nature of meaning disambiguation – the sort of disambiguation that takes place among we communicating humans is not to the same degree of precision associated with the semantic items associated with a computational device. This point emerges as a fundamental point of disagreement between the alternative view I am developing and the more classical Language of Thought view, and this is precisely the charge I shall lay against Fodor (Section 4.4.1; Denying the Compositionality, Systematicity & Productivity Desiderata; Trees and Minds). I leave my detailed analysis of the actual ambiguity of meaning until then.

I am quite happy at this point to agree, though, with the basic picture Dreyfus develops in which the use of language is a pragmatic affair and “the meaning [of some word] can always be made sufficiently unambiguous in [that] particular situation so as to get the intended result.”

Disambiguation of meaning on this view, then, is something that takes place relative to a given, particular situation; precise, unambiguous meaning is not something that stands independent of being embedded (and interpreted) relative to some particular situation. The situation itself is the larger context in which sits the word or utterance as a sub-pattern, and which provides the necessary, situated and particular disambiguation. I remain committed, however, to the further notion that meaning remains, ultimately, somewhat ambiguous, but Dreyfus’ point is slightly different.

Dreyfus argues that we are tolerant of ambiguity in the sense that our very “sense of the situation … allows us to exclude most possibilities without their ever coming up for consideration”; we can “narrow down the spectrum of possible meanings by ignoring what, out of context, would be ambiguities …[we have] ‘ambiguity tolerance’. Thus, we are capable of tolerating (read: eliminating) the sort of ambiguity associated with the potential infinity of relevant cues, since fringe consciousness provides us with some sense of the situation that limits the amount of information we are required to process. Now Dreyfus is again on familiar ground, with Wittgenstein and the like as company. It is our lived lives, our ‘forms of life’ and our developed and shared word-view that allows us to get a sense of the situation and disambiguate meaning sufficiently for communication. We are certainly not aware of whatever processing is required, so it does indeed seem relegated to fringe consciousness. Whether the required processing can be accomplished by traditional computational methods I take as an independent question for now, and acknowledge that

281 Dreyfus, 1999, pg. 108
282 Dreyfus, 1999, pg. 109
283 Dreyfus, 1999, pg. 109
the answer depends on the success of the arguments presented thus far, and in Chapter 2, but the point here is that a distinct recognitional capacity can be identified – situational recognition. We know about the world, we know about each other, and that is how we disambiguate phrases like “Mary saw a dog in the window. She wanted it.” (Such examples are artificially simple, sure, but they make the point, and they can always be constructed so as to be more ambiguous - what if we said “… She pressed her nose up against it.”). We recognize the situation, thus limiting the information that need be considered, and draw appropriate inferences.

Fine, as far as this goes, I submit, but I find unsatisfactory how Dreyfus intends to differentiate whatever is happening in fringe consciousness from the ‘sense of the situation’, and how each is meant to inform the other. He explicitly says one “presupposes the other”, and I take it that the ‘presupposing’ bit indicates that there is an informational priority in that it is fringe consciousness that is performing whatever cognitive work is required for the sense of the situation to emerge. It is unclear to me, however, how much fringe consciousness itself may be contributing directly to disambiguation (or to our sense of the situation) and how much or what cognitive work may be available to fringe consciousness due to our ‘sense of the situation’. I see no reason to believe that our ‘sense of the situation’ is not informing, or making informational available to, fringe consciousness. I do not, then, find these two definitions sufficiently distinct - although they are useful as rough psychological categories. It is my view that a proper functional analysis of C-Nets can provide a more thoroughgoing distinction (see Chapter 5), but without further analysis the role of each in meaning disambiguation, and their relation to each other, remains unclear. This point does not matter to the main argument – that there is a holistic ‘sense of the situation’, endemic in human communication, that allows for only a limited number of cues to be considered in disambiguating meaning - but I do think Dreyfus is wrong to place bets on an informational priority, given the sort of functional divisions he brings to bear.

The challenge of being able to pick out a single, relevant cue in the learning of language – or in teaching a machine to understand or translate natural language– was pointed out by Quine(1960). Dreyfus has taken Quine’s point to heart in further arguing that the problem of language learning cannot be solved by simply instilling the “sort of conditioned reflex involved in learning to associate nonsense syllables”, as there always remains some degree of ambiguity; such ambiguity that is resolved by the general recognition of some particular situation. Pointing and saying the word (say ‘chair’) does not resolve the problem by itself – are we using the proper

284 Dreyfus, 1999, pg. xix
285 Dreyfus, 1999, pg. xix
286 Dreyfus, 1999, pg. 109
name or noun? are we pointing out the colour, the shape or the size? This is exactly Quine’s point - the question remains as to what, in particular, about the chair (or rabbit or basketball game) are we pointing to? The solution to this problem of regress of reference is what Wittgenstein meant when he that said we must share a ‘form of life’, such that the child (and not the computer) “shares at least some of the goals and interests of the teacher, so that the activity at hand helps to delimit the possible reference of the words used.” We have the advantage of common evolved histories, perceptual systems, emotional systems etc. to share with our infants and set some initial conditions on our ‘sense of the situation’ – we have no such overlap with computational machines.

What we are doing then, in using our intuition in a philosophical thought experiment to judge a particular case as an instance of knowledge (or not), is recognizing that situation as exemplifying knowledge. The way in which we disambiguate the meaning is thus highly dependent on just what the ‘form of life’ it is that we bring to bear on that particular situation. It would seem no surprise then, that the intuitions of groups of people would differ - for what it is that provides the required disambiguation can be clearly identified as something that will be specific to that group. The larger context in which any particular thought experiment is framed is, of course, the particular lives lived - and the world-views commensurately generated – by the participants.

### 3.2.3 Essential/Inessential Discrimination

Here, Dreyfus looks at a class of problems in cognitive simulation known as ‘general problem solving’, and focuses on attempts to avoid the usual brute-force methods; such methods that are characterized by enumerating all possible alternatives and going through those alternatives in an orderly fashion. There is a functional similarity to ambiguity tolerance, in that it is the human ability to pick out what is essential, and to discard what is irrelevant, that is the heuristic being sought. Recall ambiguity tolerance was regarded as a distinct ability that results from the capacity to disregard the irrelevant from a potential infinite of possible relevance – as applied to the disambiguation of meaning - and here it is that capacity itself which is under scrutiny. Recognition of the relevant is not a computable function, so argues Dreyfus.

Brute force methods in problem solving are akin to the ‘counting out’ method in chess, in which one tries to go through every conceivable alternative to find the best possible move; such

---

287 Dreyfus, 1999, pg. 110
288 Dreyfus, 1999, pg. 110
methodologies are doomed due to the (by now familiar) combinatorial explosion in alternatives, and hence only work in very simple and artificial environments. We have seen there are two different sorts of failure such methods face, failure due to finite computing resources (failure *in-practice*) and failure due to the fact that the problem cannot even be defined in way that is amenable to such a method (failure *in-principle*).

Recognizing this obstacle, general problem solvers in AI attempt to find heuristics designed to generate a “grasp of the essential structure of the problem”289, which Dreyfus (following the Gestalt psychologist Wertheimer) calls “insight”290. Minsky himself has recognized the necessity of having an “ability to split or transform [the problem] into problems of a lower order of difficulty. To do this … one requires some understanding of the situation”291. The question being posed here is: What is the nature of the insight humans use to delimit problems such that they become tractable? The real problem in AI lies not in the re-structuring of problems such that the computational resources required become well-defined and tractable *per se*, but in having such re-structuring performed *without human intervention*, since it is those very heuristics that we are after; one must avoid a situation in which “this insightful restructuring of the problem is surreptitiously introduced by the programmers themselves”292.

The argument Dreyfus ultimately makes is a lack of empirical evidence in the AI sciences to produce a program that is capable of solving problems posed in a suitably general manner, and so it is not possible to provide that full argument here (it is the greater part of *What Computers Still Can’t Do* to enumerate the failures of AI to provide such an implementation). An example of a ‘cheat’, in which humans provide the necessary insight, will bring out the general line of this argument. The general point is that such a ‘cheat’ an always be found in attempts to mimic human cognition – a field known as *cognitive simulation* - and the burden of proof is shifted to those who think such a ‘cheat’ is traditionally computable.

Newell and Simon have spent considerable efforts trying to build a program that simulates what humans do in solving the artificially simple problems associated with symbolic logic (simple compared to negotiating our way around the real, messy world, for example). In their research, humans were “given problems in formal logic and a list of rules for transforming symbolic

289 Dreyfus, 1999, pg. 114
290 Dreyfus, 1999, pg. 114
291 Minsky, 1969, pg. 421
292 Dreyfus, 1999, pg. 115
expressions and asked to verbalize their attempt to solve the problems.” Attempts to simulate, in algorithmic form, the choices people made in forming solutions failed and even Newell and Simon admitted such failures “impl[ying the existence of] a mechanism (maybe a whole set of them) that is not in [the program designed to solve the problems] GPS”.

Attempts to overcome these sorts of failings inevitably involve a decision by the programmer to do something like the following: “On a purely pragmatic basis, the twelve operators that are admitted in this system of logic can be put in two classes, which we shall call ‘essential’ and ‘inessential’ operators, respectively.”

This happens, as Dreyfus points out, time and time again. The lesson from this string of failures is that “we are left with no computer theory of the fundamental first step in all problems solving: the making of the essential/inessential distinction.”

The nature of this insight into discrimination of the essential from the inessential is not independent of either fringe consciousness nor the ambiguity tolerance of the last section; it seems to be, at bottom, another way of framing the same core problem. Dreyfus acknowledges the relation to fringe consciousness in that playing chess, in which massive unconscious work is done in eliminating the inessential, is really just another example of problem solving. Players are getting at the core of the problem, and displaying the relevant kind of insight and problem-space reduction, by making use of fringe consciousness to allow them to zero-in on a finite set of relevant possible moves. I take it that the learning process alluded to in the previous section, and the tolerance of the infant of an irreducible ambiguity of reference, is also an instance of the essential/inessential distinction; in that case, the infant gets at what is essential by sharing our ‘form of life’.

I also take it that essential/inessential discrimination occurs right across the epistemic continuum. What makes a particular maple tree a maple tree is not any of an infinite number of possible similarities or dissimilarities (the movement of the leaves in the wind, the pattern of veins on the leaf, the number of birds that have landed on it in the last three hours, etc.) but those similarities that are essential. The same goes for a pick-and-roll, a weak king side, or an instance of ‘graceful’ or ‘knowledge’.

3.2.4 Perspicuous Grouping and the Pattern-Rec Stance

**Perspicuous grouping** is the sort of pattern recognition ability that is a culmination of this group of

---

293 Dreyfus, 1999, pg. 113
294 Newel & Simon, as quoted in Dreyfus, 1999, pg. 114
295 Dreyfus, 1999, pg. 115
cognitive abilities. A parallel metaphysics of pattern was developed in the previous chapter; objects and events, or ‘intermediate regularities’ as Dennett would call them, can exhibit an open-ended complexity such that they do not need to have any particular set of traits in common across the group. Such regularities include the kind of similarity associated with Wittgenstein’s family resemblances, and membership in these groups is open-ended in that new members can be added on a pragmatic case-by-case basis; “Pattern recognition in this domain is based on recognition of the generic, or of the typical, by means of a paradigm case.”297 Here Dreyfus argues that our ability to group such regularities together is dependent on the three cognitive abilities listed in this section. What humans do, on an everyday basis, is to recognize “patterns as complex as artistic styles and the human face”298 and these “reveal a loose sort of resemblance which seems to require a special combination of insight, fringe consciousness, and ambiguity tolerance beyond the reach of digital machines”299.

This recognitional ability is then, quite distinct from the sorts of pattern-recognition performed by computers, which must “recognize all patterns in terms of a list of specific traits”300, and the point I am making here is that, taken together, these abilities allow for a type of pattern recognition to take place, which Dreyfus has called perspicuous grouping. Perspicuous grouping is really another way of saying that one has the ability to notice relevant similarities, and is – for Dreyfus – one of the more important ways in which human behaviour differentiates itself from the potential behaviour that could be exhibited by a rule-following computer. Rather than following rules, the abilities outlined in this section lead to a situation in which humans “seem to be using global perceptual organization, making pragmatic distinctions between essential and inessential operations, appealing to paradigm cases, and using a shared sense of the situation to get their meanings across”301.

These abilities act as a set of computing primitives - defined in functional terms – which, when taken together, act as an operational definition of a pattern-recognizing system. I am simplifying this group of abilities into the expression similarity-for-free. Similarity-for-free represents a computing primitive different in kind from the sorts of computing primitives associated with a Turing machine, and overcome the limits associated with the Level II – III distinction as it applies to the Church-Turing thesis.

296 Dreyfus, 1999, pg. 117
297 Dreyfus, 1999, pg. 294
298 Dreyfus, 1999, pg. 120
299 Dreyfus, 1999, pg. 120
300 Dreyfus, 1999, pg. 120
Adopting the pattern-rec stance toward mind is to take these recognitional abilities as operational primitives of the system, to take them as *what the system must be taken to be able to do* if it is to be classed as such a system, and to take the ability to recognize the relevant patterns as granted if the system has such abilities.

In chapter 5, I shall further develop this notion in terms of how C-Nets are capable of implementing this sort of system.

301 Dreyfus, 1999, pg. 286
Chapter Four

4. CONCEPTUAL ATOMISM – FODOR ON CONCEPTS

Here, I present a classical view of concepts, one most clearly articulated by Fodor, that is distinct from the view I will present in the next chapter, most particularly in view of what concepts are, and what it means to possess a concept. Pragmatic conceptualism characterizes concepts primarily as epistemic capacities, and on Fodor’s view they are to be regarded, above all, as mental entities. Fodor has no stomach for this very distinction, as he thinks it betrays a previous commitment to dubious metaphysics:

... when philosophers take a strong line on a methodological issue there’s almost sure to be a metaphysical subtext. ... Suffice it, for now, that the thesis that concepts are mental particulars is intended to imply that having a concept is constituted by having a mental particular, and hence to exclude the thesis that having a concept is, in any interesting sense, constituted by having mental traits or capacities. ... the methodological doctrine that concept possession is logically prior to concept individuation frequently manifests a preference for an ontology of mental dispositions rather than an ontology of mental particulars.\(^3\)

Fodor’s view of concepts is deeply entwined with his view of thought as symbolic computation, and my arguments against his theory of concepts here is supported mainly by way of denying the view of mind as Turing machine. Issues of metaphysics have been previously discussed in Chapter 2; such issues are put aside here to make way for arguments of a more psychological nature. Although I disagree with the basic premises of Fodor’s view of cognition, in a sympathetic reading of the ‘atomism’ portion of informational semantics – a kind of relation between mind and world – an interestingly similar way of viewing concepts as reliant on intuition as pattern-recognition will be articulated. Indeed, leaving aside issues of how the brain computes, and focussing instead on how the mind gains semantic access, Fodor’s conceptual atomism and my Real-H-Intuition make strange, but compatible, bed-fellows.

My strategy of getting at Fodor’s theory of concepts will be the following. First, I outline what

\(^3\) Fodor, 1998, pp. 4-5
motivates him in the first place to develop such a theory. That motivation turns on defending a Representational Theory of Mind (RTM), which itself turns on a list of what I call psychological desiderata, a set of mental performance criteria that need explaining - the psychological explananda. Those explananda then, in turn, rely partly on a mix of quasi-empirical observation but mainly on a preference that psychology, as a science, reflect the nomic character of the hard sciences. The derivation of Fodor’s motivation is presented (but not criticized) in Motivation and the Psychological Explananda (4.1). I then extensively explore what Fodor means by a Representational Theory of Mind (4.2). This section will be my focus, since it is what motivates the larger Fodorian picture. The explanatory framework set out, I then characterize Conceptual Atomism (4.3) itself. My arguments against Fodor’s view are presented in Fodor: Analysis (4.4) and, finally, I develop a Sympathetic Reading of Atomism and Informational Semantics (4.5).
4.1 Motivation and the Psychological Explananda

4.1.1 Fodor as Coherentist

Before characterizing the theory itself, it is worth noting Fodor’s motivation in developing an atomistic theory in the first place. Fodor’s atomistic theory of concepts is *coherentist* in nature, I claim. It is coherentist since it is derived as a means to support a secondary explanatory framework, (the RTM), one that claims as its primary strength a unique ability to account for a wide range of psychological phenomena. Several of those psychological phenomena, in turn, are derivative of a particular view of psychology as a kind of law-like science. It is, ultimately, a desire to see all of those psychological desiderata explained in one, fell, coherent swoop that draws Fodor into wanting an RTM, and the theory of concepts follows from that RTM.

Fodor is thus primarily concerned with adopting (and thus defending) a Representational Theory of Mind; he has a long history of defending RTMs, he has, indeed, “... grown old writing books defending RTMs”\(^{303}\). Once a RTM has been adopted (and independently argued for) then an atomistic theory of concepts is, for Fodor, the only theory that will fit the bill. Thus, defending RTMs is the primary argument Fodor needs to sustain, for the atomist theory is an extension of (and thus reliant on) that argument. Fodor;

“.. if you are going to run a representationalist/computational theory of mind (that is, any version of RTM ...) you will need a theory of concepts.
And:
.. none of the theories of concepts that are currently taken at all seriously either in cognitive science or in philosophy can conceivably fit the bill”\(^{304}\)

I take note one doesn’t need to run an RTM in order to require a theory of concepts; presumably, anyone ‘running a theory of mind’ *at all* would need one.

For Fodor, RTMs are unique in their being able to provide an explanatory framework that covers a huge amount of what he argues a theory of mind *ought* to explain; RTMs explain a good deal of mental desiderata. While I agree that RTM does indeed explain the desiderata as defined by

\(^{303}\) Fodor, 1998, pg. 1
\(^{304}\) Fodor, 1998, pg. 23
Fodor, and does so quite well, I argue in the final section that the desiderata themselves are misconstrued. While Fodor provides theoretical considerations for the adoption of the desiderata themselves, those considerations are really just a preference for a particular explanatory framework, namely, that of equating the form of psychological explanation with the form of scientific explanation (what I mean by this claim is explained in the next section). Without independent empirical or theoretical justification for that preference, he is just using a kind of global intuition to try and build the most coherent story. He has, by insisting on sticking (to the bitter end) with what he takes to be a good, coherent theoretical story, committed himself to a characterization of human cognition that is just not - as I argue in the Analysis section - empirically credible.

There is a kind of two-way street of influence here, which is why I think it fair to characterize Fodor’s story as strongly coherentist. The five desiderata outlined below are best explained by an RTM, but at the same time, an RTM also requires those desiderata. I take Fodor’s direction of logical priority to be different for the first two - causality and publicity - from the priority of the last three - compositionality, systematicity and productivity. Causality and publicity appear to have an independent (though false) argument to support them. The remaining three - productivity, systematicity and compositionality - seem to be as much an outcome (or output) of RTM as they are independently supported psychological considerations (a position I argue for in Analysis). That Fodor sees them as empirically credible psychological phenomena, though - that they are meant to reflect the way in which we actually think - he makes abundantly clear.

Whatever the order of priority, however, for Fodor the relationship is conceptual as there is very little, if any, empirical evidence given to directly support the desiderata themselves. Without such evidence, one cannot help but be suspicious that those capacities are an idealized view of human cognition, and not reflective of actual, instantiated, human thought.

4.1.2 The Desiderata; Causality & Publicity

An independent argument to support RTM is roughly that an intentional explanation, to be proper science, must provide for law-like generalizations; “Psychological explanation is typically nomic

---

305 I differentiate my use of the term ‘coherentist’ here from that normally associated with epistemic concerns (to be a coherentist about knowledge) - what I mean by the term is that Fodor is motivated by considerations of coherence in building his theoretical structure, the coherence applies to the structure of his overall theory, not to knowledge itself.
and is intentional through and through"\textsuperscript{306}. Note that ‘nomic’ here requires the form of laws found in science – there are predicates defining the objects over which the law applies, and a relationship of necessity between the antecedent and consequent; “if there are no intentional laws, then you can’t make a science out of intentional explanations [emphasis mine]”\textsuperscript{307} The emphasis here is placed on the existence of clear and well-defined predicates, and clear and well-defined necessity relations, each unencumbered by any required accompaniment by numerous and weighty ceteris paribus conditions (a topic about which I have much to say presently); “RTM can maybe provide the mechanism whereby satisfying the antecedent of an intentional law necessitates the satisfaction of its consequent.”\textsuperscript{308}

What are the two desiderata such that intentional explanation falls under a nomic rubric?

1. Causality: Fodor wants mental particulars to be part of a causal and nomic explanation of human behaviour. If psychology is to be a science, then that which psychology explains had better fall under nomic (law-like) generalizations; “The ‘.. mental particulars ...satisfy whatever ontological conditions have to be met by things that function as mental causes and effects.”\textsuperscript{309} and “The laws the psychological explanations invoke typically express causal relations among mental states that are specified under intentional description”\textsuperscript{310}. What this means is that when I go to the store because I want a carrot, my action is caused by a token mental state corresponding to WANTING A CARROT. Fodor has a quite narrow view of what constitutes a science here, and what notion of causation may support a science - only law-like generalizations may apply.

2. Publicity: Concepts are “the sorts of things that lots of people can, and do, share”\textsuperscript{311} Mental tokens fall under types, and it is that single concepts fall under the same type from person to person, and in a single person over time, that provides for the publicity requirement; “to say that two people share a concept (ie that they have literally the same concept) is thus to say that they have tokens of literally the same concept type”\textsuperscript{312} . I take it that this applies to all concepts, from those that apply to natural kinds to those that are highly abstract or arguably somewhat subjective, such as KNOWLEDGE or GRACEFUL. People who exist in dissimilar perceptive modes or have quite different cognitive capacities also share the same mental types;

\textsuperscript{306} Fodor, 1998, pg. 7  
\textsuperscript{307} Fodor, 1998, pg. 7  
\textsuperscript{308} Fodor, 1998, pg. 13  
\textsuperscript{309} Fodor, 1998, pg. 23  
\textsuperscript{310} Fodor, 1998, pg. 7  
\textsuperscript{311} Fodor, 1998, pg. 28
...people who live in very different cultures and/or at very different times (me and Aristotle, for example) both have the concept FOOD; and that people who have had very different kinds of learning experiences (me and Helen Keller, for example) both have the concept TREE; and that people with very different amounts of knowledge (me and a four-year-old, for example) both have the concept HOUSE. And so forth.\textsuperscript{313}

As we will see, concepts get their content from the world, and possess that content by virtue of their syntactic form - thus, since (for Fodor) a tree is a tree is a tree, it follows that the tokened mental concept TREE will be the same across individuals. The world constrains concept content, the world is the same for each of us (a tree is a tree is a tree), concept content defines the concept token; thus - those tokens are identical. Meaning, for Fodor, is information and information is something about the world; “the content dog is something about its tokenings being caused by dogs”\textsuperscript{314} - I have more to say about this presently. Fodor’s commitment to this view is deep enough that it has a logical priority over any method that defines concepts in such a way that they can turn out to be distinct;

... if a theory of an experimental procedure distinguishes between my concept DOG and Aristotle’s, or between my concept TRIANGLE and Einstein’s, or between my concept TREE and Helen Keller’s, etc. that is a very strong reason to doubt that the theory has got it right about concept individuation or that the experimental procedure is really a measure of concept possession\textsuperscript{315}

Fodor believes he has independent grounds for defending this view, that it relies on a prior ontological commitments - a tree being a tree being a tree. Essentially though, it is his commitment to RTM that drives these ontological considerations, “... I think there are good grounds for taking a firm line on this issue [against conceptual relativism]. Certainly RTM is required to.”\textsuperscript{316} And, one may fairly ask, why is RTM so attractive? “... the extent to which an RTM can achieve generality in the explanations it proposes depends on the extent to which mental contents are supposed to be

\textsuperscript{312} Fodor, 1998, pg. 28  
\textsuperscript{313} Fodor, 1998, pg. 29  
\textsuperscript{314} Fodor, 1998, pg. 12  
\textsuperscript{315} Fodor, 1998, pg. 29  
\textsuperscript{316} Fodor, 1998, pg 29
shared"³¹⁷. The ‘generality’ requirement here, is reflective of science-like intentional explanation – the generality is such that the predicates of nomic psychological explanation are identical across people, places and times.

So publicity and causality are required by RTM, which is a ‘good thing’ because it keeps intentional explanations law-like; “Prima facie, it would appear that any very thoroughgoing conceptual relativism would preclude intentional generalizations with any very serious explanatory power. This holds in spades if, as seems likely, a coherent conceptual relativist has to claim that conceptual identity can’t be maintained even across time slices of the same individual”³¹⁸

Just so - in my view, so much the worse for a science-like (even physics-like), mechanical, covering-law intentional explanations. At this point I simply note that there are other possibilities, notably non-nomic intentional explanations in which the intentional realm is autonomous, but falls under intentional explanation. Davidson famously carries such an argument in advocating ‘anomalous monism’, in arguing against type-type identity and for token-token identity and the autonomy of the mental realm; “The thesis is rather that the mental is nomologically irreducible: there may be true general statements that have the logical form of a law; but they are not lawlike”³¹⁹. Fodor is completely dismissive of this possibility, denying he even understands what it means; “I find the arguments that there are no intentional laws very hard to follow ... [and] ... I find all that realm talk very hard to follow too.”³²⁰ Methinks the man overly shy of his mental capacities.

A mere assertion that an explanatory scheme cannot be made of non law-like descriptions seems untenable, and requires support. Indeed, Cartwright³²¹ has made much of the non law-like nature of much of science itself. Without considerable argument, it really just seems a deep intuition of Fodor’s that proper explanations require the law-like stature of something like physics. I argue this point further in Analysis, and emphasize, analogous to Cartwright’s view of science itself, that it is the implicit ceteris paribus clauses that do most of the real work in allowing for the kind of generalizations we look for in good explanations. Intentional explanation is different from scientific explanation, but no less an explanation for so being.

³¹⁷ Fodor, 1998, pg. 29
³¹⁸ Fodor, 1998, pg. 29
³¹⁹ Davidson, 1980, pg. 216
³²⁰ Fodor, 1998, pg. 7
³²¹ Cartwright, 1999, 1983
4.1.2.a Interlude: Publicity, Similarity and Identity

Fodor’s publicity requirement is strong; identity of concepts is required for them to be shared, mere similarity will not do. Fodor argues that it is not logically possible to claim content similarity without requiring an account of content identity. Thus, for two people to share, even remotely, the concepts that are the content of some communicative exchange, they must share identical concepts\textsuperscript{322}.

His target are those conceptual relativists (like myself) who want to do away with identity and go with similarity: Chapter 3 of this thesis is one of those “at least half a zillion [places] in the philosophical literature, where the reader is assured that some or all of his semantical troubles will vanish quite away if only he will abandon the rigid and reactionary notion of content identity in favour of the liberal and laid-back notion of content similarity”\textsuperscript{323}.

Here, the idea Fodor is attacking is the idea that “intentional explanation can, after all, be preserved without supposing that belief contents are often - or even ever - literally public. The idea is that a robust notion of content similarity would do just as well as a robust notion of content identity”\textsuperscript{324}. Quite so - however, Fodor questions whether one could do so without begging the question, if two demands are to be met; “One the one hand, such a notion must be robust in the sense that it preserves intentional explanations pretty generally; on the other hand, it must do so without itself presupposing a robust notion of content identity”\textsuperscript{325}.

I accept the ‘pretty generally’ demand (but without accepting the law-like form of such explanations) but I claim the notion of similarity developed in Chapter 3 does indeed generate a notion of similarity without presupposing identity. The argument is empirical, not logical, in that I have claimed that similarity can be defined as just that which is taken to be similar by neural networks. I take comfort that in a footnote Fodor writes, “Why not take content similarity as primitive and stop trying to construe it”\textsuperscript{326} - this is precisely what Chapter 3 purported to

\textsuperscript{322} I note here that there are as many ways to get publicity as there are newspapers in Europe. The requirement for publicity, if one grants it, does not have to rest in mental tokens in the heads of the people – it can rest in public language, and other various and sundry items in view of the good people. Thus, even if I grant the necessity of identity for similarity – which as made manifestly clear in Chapter 3 I do not (although for communication we do need shared public symbols which are taken as identical, like words) – identity itself need not be bound up in shared mental tokens.

\textsuperscript{323} Fodor, 1998, pg. 34
\textsuperscript{324} Fodor, 1998, pg. 30
\textsuperscript{325} Fodor, 1998, pg. 30
\textsuperscript{326} Fodor, 1998, pg. 32, footnote
accomplish – similarity as a basic property that is *demonstrated* by neural networks. That said, Fodor’s challenge is as follows.

Fodor’s target are those who claim some sort of similarity metric. As far as Fodor can see, any such metric take on board some variant of “*partial overlap* of beliefs”\(^{327}\). If we both have different, but similar, concepts of George W. Bush, then presumably we would agree about various properties of the man; that he is the President, that he wears suits a lot, that he smirks, and so on. We also disagree about some properties; that he is fit to govern, for example. For Fodor, the similarity of our concepts of GWB is governed by “some (presumably weighted) function of the number of propositions about him that we both believe”\(^{328}\). The question, for Fodor, is whether the shared beliefs are *literally* shared. If they are, then we need a robust notion of identity, if they are not, then “his account of content similarity begs the very question it was supposed to answer”\(^{329}\) - which was to explain similar but not identical contents, without presupposing similar but not identical contents.

The trouble is that “all of the obvious construals of *similarity of beliefs* ... take it to involve *partial overlap* of beliefs”\(^{330}\). The beliefs that overlap are either similar or identical. If they are similar, we’ve just pushed the problem back. If they are identical, the publicity is presupposed. Fodor goes through various notions of similarity to see if the accounts get away with similarity without identity. They do not.

Subjective probabilities assigned to the various properties associated with a concept, for example, also presuppose a prior notion of belief identity, as “there are lots of thoughts of which our respective PRESIDENTs are constituents that we literally share”\(^{331}\). The difference is in *strengths* of beliefs, not belief *content* - the content itself is still taken to be identical. The same holds if concepts are taken to be sets of features; the “atomic feature assignments must themselves be construed as literal”\(^{332}\). Connectionist multi-dimensional vector spaces don’t fare any better, as Fodor argues that the “dimensions which express degrees of *the very same property*”\(^{333}\).

I argue in the next chapter that similarity can be regarded as basic for a neural network, basic to its

---

327 Fodor, 1998, pg. 32
328 Fodor, 1998, pg. 31
329 Fodor, 1998, pg. 31
330 Fodor, 1998, pg. 32
331 Fodor, 1998, pp. 32-33
332 Fodor, 1998, pg. 33
computing in a way that syntax and computation-cum-causality is basic to a Turing machine. I also argued that the dimensions along which we interpret these systems is just that - an interpretation, a version of the ‘canned intentionality’ that Searle worried about - and not basic to the system itself. I leave the reader to judge the success of those arguments that one does indeed get similarity for free in a neural network, without presuppositions of identity.

That said, it seems that one can simply give up and take on board the publicity requirement, without assuming metaphysical identity. I take up this argument in the final section of this chapter, so I will only highlight the form of the response. If publicity is to be the measure of shared content, so be it, but take note that publicity can take many forms - in particular, it needs only be a method of triangulating oneself enough for public agreement. Publicity may be just that – an agreed upon method of solving inter-subjective disagreement. Fodor begs the question as to the nature of the identity conditions that may satisfy a publicity requirement in demanding that they be some prior, existent notion of identity. Perhaps there is no such thing. Perhaps the best we can do is engage in public discourse, refer to experts and empirical evidence when we disagree what our words and concepts might mean, and call the publicity requirement a ‘discursive agreement’. So much for nomic intentional generalizations, but perhaps that’s an acceptable cost of a weaker notion of publicity.

So, I will argue that one can accept the publicity requirement, without accepting the need for metaphysical identity across concepts. I have also argued that one can accept the empirical existence of a similarity metric, without presupposing identity conditions for that which the metric measures.

4.1.3 The Desiderata: Compositionality, Systematicity & Productivity

The next three desiderata are regarded as non-negotiable characteristics of concepts because, for Fodor, they follow on from a RTM; they are “... constraints on concepts that follow on from the architecture of RTMs ...”\(^{334}\). However, they are also meant to have some independent intuitive support, a bit of reflection upon one’s own thinking is meant to make clear that they are a psychological reality that is best explained by RTM. It is, in Fodor’s view “...a point of empirical fact ...”\(^{335}\), for example, that there are no minds that do not exhibit systematicity. These are presented then, as psychological facts to be explained by RTM, even though this can be regarded as

\(^{333}\) Fodor, 1998, pg. 34
\(^{334}\) Fodor, 1998, pg. 23
\(^{335}\) Fodor, 1998, pg. 26
a reversal of conceptual priority. I emphasize that I take this to be a coherentist account, driven by
an overriding need to subsume mental life into law-like intentional generalizations.

3. Compositionality: There are two distinct, but related, psychological capacities that are derived
from this psychological characteristic. The basic idea is that, from a finite set of primitive discrete
elements or concepts, one can produce an infinite variety of thoughts (productivity) that exhibit an
internal symmetry (systematicity) and the meaning of the larger, productive structures is literally
composed of the meanings of the primitive bits of which it is comprised (compositionality). Fodor
has explicated each of these capacities independently, but has in the most recent literature, nicely
folded them into compositionality, and the capacities it explains:

   “Compositionality: concepts are the constituents of thoughts, and, in indefinitely
   many cases, of one another. Mental representations inherit their contents from the
   contents of their constituents”

Although this is a statement about concepts themselves, this is meant to capture some characteristics
of human thought (or belief, or other intentional states; Fodor is using “‘thoughts’ as [his] cover
term for the mental representations which ... express the propositions that are the objects of
propositional attitudes”). That is, one is meant to have the capacity to have beliefs, or express
thoughts, of a form such that the content of a complex thought is expressed by the content of the
constituents of that thought. The meaning of a complex thought is linearly isomorphic on the
meanings of the constituent thoughts, or, less pedantically, the meaning of complex thoughts is
simply additive from the meaning of the constituent elements of that thought. That characteristic of
thoughts - compositionality - is meant to explain two capacities of human thought: productivity and
systematicity.

Note that this notion of compositionality is highly analogous to the structure found in natural
language. Indeed, it becomes an issue of great significance as to which has metaphysical or
epistemic priority - language or thought. To preview, on Fodor’s view, it is thought that provides
the structure for language. Opposing views suggest otherwise, that it is language as a kind of
external tool, that provides for a structuring of the mind. Others, myself included, take the position
that the truth of the matter lies in some murky area between the two.

336 Fodor, 1998, pg. 25
337 Fodor, 1998, pg. 25
4. **Productivity**: Compositionality of thought is meant to explain our capacity to generate, or entertain, an infinite number of thoughts; “There are infinitely many thoughts because, though each mental representation is constructed by the application of a finite number of operations to a finite basis of primitive concepts, there is no upper bound to how many times such operations may apply in the course of a construction [emphasis mine].” Details of how the thoughts are generated aside (for now) the point to note is that Fodor is claiming that we can entertain, presumably, thoughts of infinite size and complexity (short-term memory considerations aside).

5. **Systematicity**: Another of the capacities that is explained by compositionality is meant to be the natural, internal systematicity exhibited by thought; “Beliefs are systematic in that the ability to entertain any one of them implies the ability to entertain many others that are related to it in content.” My capacity to believe that ‘Fred loves Dave’ is meant to ensure that I have the capacity to think that ‘Dave loves Fred’. I take note that these paradigmatic examples of systematicity can also be seen as a sort of internal symmetry; an ability to have a thought like ‘The stove is beside the fridge that is broken’ is presumably meant to guarantee the ability to entertain a thought like ‘The fridge that is broken is beside the stove’. Again, thought exhibits the kind of systematicity found in natural language, which inherits its structure from the compositional nature of concepts. It is the nature of concepts that ensures that our thinking exhibits systematicity. Combined with productivity, presumably the kinds of complicated internal symmetries one could construct with ever more complicated sentences, is boundless.

Fodor admits that it is not a conceptual necessity that such is the case, but claims that it is an empirical fact about human cognition; “It appears, for example, to be conceptually possible that there should be a mind that is able to grasp the proposition that Mary loves John but not able to grasp the proposition that John loves Mary. But, in point of empirical fact, it appears that there are no such minds. This sort of symmetry of cognitive capacities is a ubiquitous feature of mental life.” I must say, I’m glad Fodor places such emphasis on the empirical grounds for such a claim, as I rely on equally empirical grounds in refuting these desiderata. For now, though, let us note, that it is not language itself that Fodor is claiming exhibits this systematicity, but thought itself, minds themselves.

I emphasize in closing that the above cognitive capacities are not presented by Fodor as some

---

338 Fodor, 1998, pg. 27
339 Fodor, 1998, pg. 26
340 Fodor, 1998, pg. 26
idealized view of human cognition, or as properties of natural language as a construct external to the mind, but rather they are meant to be a description of actual human cognition. It is not the case that Fodor is claiming that we are - roughly - productive, systematic, compositional in our thinking, that roughly our concepts are the exact same things, and that roughly our concept tokens fall under identical types across people, cultures and time. He claims that we are precisely productive, systematic and compositional in our thinking. He claims that we mean exactly the same thing when we talk of some concept in using the same word. He claims that our concept tokens, that which causes our behaviour to fall under intentional description, are exactly the same tokens across individuals, time and space; this precision is that which allows intentional description for him to exhibit the same law-like precision as the laws of science.

It is a kind of crystalline structure that he purports is reflective of our cognitive capacities; our thinking is precisely productive, systematic, compositional, causal and public. I will have more to say about what I mean by ‘crystalline’ (which, following de Sousa\textsuperscript{341}, is related to ‘digitality’) - and why I think Fodor’s claims about human thought can be characterized in this way - in Analysis. To preview - I argue that Fodor’s account is coherentist\textsuperscript{342}, in the sense that it is not clear which has explanatory priority – the psychological \textit{desiderata} (as empirical facts to be explained) or the RTM (as a prior commitment that requires the \textit{desiderata}). The RTM is clearly meant to explain the desiderata, but at the same time, it seems that the desiderata have been overstated in the degree of digitality they display, in order to \textit{accommodate} the RTM. So I turn now to the question – what is it that constitutes an RTM?

\textsuperscript{341} De Sousa, 1991
\textsuperscript{342} Again, the coherence to which I am referring arises from considerations of the structure of his overall theory – not coherence as it is sometimes used to refer to the structure of knowledge itself.
4.2 Representational Theory of Mind (RTM)

The clearest expression of a classical RTM is a Turing machine, and Fodor is indeed quite enamoured of the idea that the mind *is* a Turing machine; “I propose to swallow the Turing story whole and proceed.”\(^{343}\) It is not so much that the mind *can* represent (I don’t know if there are many people around today who would deny that sort of claim – other than perhaps Churchland), it’s that representation and computation *constitute* thought, or what the mind does. For Fodor, the mind *as* a Turing machine operates on little syntactically defined symbols (representations), churning out computations (causal relations among symbols) that treat those syntactic bits as basic concepts, producing thought. The mind is, at bottom, a set of representations and operations on those representations. Representing is *basic* to all thought, it precedes thought, and “computations are those causal relations among symbols which reliably respect semantic properties of the relata.”\(^{344}\).

What does it mean to be a Turing machine? Fodor presents RTM as a series of five theses.

4.2.1 Nomic Explanation

As outlined in the previous section, a thesis Fodor holds very dear to his heart is that “*Psychological explanation is typically nomic and is intentional through and through* [emphasis his]”\(^{345}\). A description about behaviour in everyday language, like ‘David went to the store because he wanted some carrots [and believed that there are carrots at the store]’ is a typical intentional explanation. For Fodor, to attribute such a description is to be very literal about what it is in David’s head that allows us to treat that everyday description as an intentional explanation, and enables us to view that explanation as universal in scope, in the very same way that scientific predicates and the laws under which they are subsumed are universal in scope. That is, there are “causal relations among *mental states that are specified under intentional description*”\(^{346}\).

An intentional description is law-like, and there is literally tokened in David’s head little mental bits that correspond to ‘carrot’, ‘store’, and the like that are the physical cause of David’s actions. There is nothing *ceteris paribus* here; the intentional description in all its generality holds as law,

\(^{343}\) Fodor, 1998, pg. 10
\(^{344}\) Fodor, 1998, pg. 10
\(^{345}\) Fodor, 1998, pg. 7
\(^{346}\) Fodor, 1998, pg .7
and the nature of the syntactic bits that act as predicates are identical, across people, contexts and time. Little carrots, big carrots, a department store, a corner store, David’s head, David’s father’s head; none of this matters – the syntactic bits remain identical.

4.2.2 Mental Representations Bear Original Intentional Content

Intentional content derives from mental entities; “‘Mental representations’ are the primitive bearers of intentional content.” By ‘primitive’, Fodor means that the intentionality of language, the means by which language refers to some state of affairs, is derivative of the intentionality of the propositional attitudes (those intentional descriptions we use to explain behaviour, see above), and the intentionality of propositional states is derivative of the mental entities underpinning those propositional states (that which serves to act as the physical cause in the intentional explanation). It is the mental representations that provide, both explanatorily and ontologically, the intentionality of propositional attitudes, and the language in which those propositions are expressed. Fodor:

Both ontologically and in order of explanation, the intentionality of the propositional attitudes is prior to the intentionality of natural languages; and, both ontologically and in order of explanation, the intentionality of mental representations is prior to the intentionality of propositional attitudes.

Mental representations, those things that we have seen play a role in intentional explanation and which serve to act as the physical items that cause the intentional behaviour, are the intentional bedrock upon which sits natural language. Any event described by a propositional attitude (a belief, say), with content P, corresponds to some other event that “consists of the creature’s being related, in a characteristic way, to a token mental representation that has the content P”. Fodor is perfectly happy with the old ‘belief box’ analogy; “For each episode of believing that P, there is a corresponding episode of having, ‘in one’s belief box’, a mental representation which means that P.”

The main point here can be put thus: in order of intentional priority it’s mental representation first, thought second, propositional attitude third, and language last. This thesis “contemplates no locus

---

347 Fodor, 1998, pg. 7
348 Fodor, 1998, pg. 7
349 Fodor, 1998, pg. 8
350 Fodor, 1998, pg. 8
of original intentionality except the contents of mental representations.”

For Fodor, learning a language does not consist so much in learning the meanings of the words (as that is derivative upon mental representations) but “learning to associate its sentences with the corresponding thoughts.” There is, then, the “metaphysical possibility of thought without language” as well as the “metaphysical possibility of mental representation without thought.”

4.2.3 Thinking Equals Computation

Here, Fodor acknowledges an intellectual lineage back to Hume, for whom mental states are some sort of relation to a mental representation, and who “taught that mental processes (including, paradigmatically, thinking) are causal relations among mental representations”. Mental states are static, they are a relation to a mental representation, which is itself an entity whose semantic properties are defined syntactically. Those causal relations that act on/between those syntactically defined representations are, for Fodor, computation - a dynamic process. One of the important properties of a Turing machine is that those causal relations, in acting solely on the syntactic properties of the representation (of which the semantic properties are derived), are truth-preserving.

Fodor; “Turing’s account of thought-as-computation showed us how to specify causal relations among mental symbols that are reliably truth-preserving.” Computation is then “some kind of content-respecting causal relation among symbols”. Symbols, or mental representations, “have constituent (part/whole) structure, and many mental processes are sensitive to the constituent structure of the mental representations they apply to.” Mental representations, then, can be atomic but can also be molecular. The meaning of the whole is the sum of the meanings of its parts.

It is important for Fodor that the notion of symbol “doesn't itself presuppose the notion of computation”. Computation has been introduced as “some kind of content-respecting causal relation among symbols”. Here, there lurks a danger of circularity. Fodor wants to ensure that he doesn’t “... need the notion of a computation to explain what it is for something to have

---

351 Fodor, 1998, pg. 8
352 Fodor, 1998, pg. 9
353 Fodor, 1998, pg. 9
354 Fodor, 1998, pg. 9
355 Fodor, 1998, pg. 10
356 Fodor, 1998, pg. 10
357 Fodor, 1998, pg. 11
358 Fodor, 1998, pg. 11
359 Fodor, 1998, pg. 11

139
semantic properties.” It is by having semantic properties by virtue of which something is a symbol, and not merely physical syntax. Computation is not causation, for Fodor. Thus, Fodor must fill in his promissory note that “... the account of the semantics of mental representations that [his] version of RTM endorses, unlike the account of thinking that it endorses, in indeed non-computational.”

Here, computation has been presented as an operation on a representation, which in turn is a symbol - it has semantic content. Representation - implying semantic content - is bound up with computation whereas syntax is bound up with causation. A syntactic engine without interpretation does not contain symbols, it is by bestowing upon them the status of symbol that Fodor gets away with the very notion of content. Pylyshyn has previously noted that any computing device requires a basic set of physical resources which he calls “basic computational resources”, a set of basic causal relations, that are assumed to be in existence prior to being able to talk about computational processes on syntactically-defined symbols. Here, we separate causation - the laws of nature by which the process works - and computation. Since representation is here bound up with computation, and since the representation is a syntactic expression, Fodor will need to separate the semantics of representation - that which makes the syntactic token have content and thus attain the status of symbol - from computation to get away with this definition. That task is what is meant to be accomplished by informational semantics, which is outlined in the next section.

There is a distinction to be made here. Fodor allows that computation and symbol can be inter-defined, it is only when someone is pushing the line that “computation [is] to be part of a naturalistic psychology; viz. part of a programme of metaphysical reduction” that the problem arises. What he wants to avoid here, presumably, is Searle’s complaint that any symbol-denoting status, that is, any semantic content, that is bestowed upon syntactic structures in machines that are not conscious, is a kind of canned semantics. It’s only by virtue of the fact that the ‘symbols’ are interpreted that they have content, and - for Searle, at least - it is a conscious designer/operator that is providing the interpretation. By naturalistic psychology, Fodor is grounding all such interpretation (conscious or not) in the notion of representation.

Thinking, then, consists of the causal processes that take place among tokened symbols in the head

---

360 Fodor, 1998, pg. 11
361 Fodor, 1998, pg. 11
362 Fodor, 1998, pg. 11
363 Pylyshyn, 1985, pg. 259
364 Fodor, 1998, pg. 11, footnote
- and we need a non-computational account of semantics of the representations to avoid Searle’s charge. Those causal processes, which preserve the semantic/syntactic structure of the symbols (and thus are truth-preserving), are computation. Thus, thinking is computation.

Fodor & Pylyshyn; “It would not be unreasonable to describe classical cognitive science as an extended attempt to apply the methods of proof theory to the modelling of thought (and similarly of whatever other mental processes are plausibly viewed as involving inferences)”. Here we have the baldest claim of the Leibnizian project of a computable universe (that which provides for conceptual content), and a mind that can compute it (that which manipulates that content).

4.2.4 Informational Semantics

What Fodor needs to accomplish, then, is a means of de-coupling computation from semantics. That which provides for a syntactic structure to obtain semantic status needs to be something other than the way in which that syntactic structure is defined as a part of the computational process that treats it as a symbol, rather than a mere physical token. The syntactic structure is operated on by physical law; that physical law can be regarded as computation if the syntactic structure is elevated to the status symbol (in a computational process), and the syntactic structure can be regarded as symbol if the physical law is elevated to the status of computation (over that symbol). However, as we have seen, physical laws and syntactic tokens cannot mutually boot-strap each other to the status of computation and symbol on fear of circularity. Fodor needs an extra ingredient to bestow semantic status to the syntactic tokens, and that extra ingredient is informational semantics.

By informational semantics, Fodor means that semantic content is exhausted by a particular kind of mind-world relation; “Pure informational semantics allows me to hold that one’s inferential dispositions determine the content of one’s concepts because it says that content is constituted, exhaustively, by symbol-world relations”. As we will see, the stronger atomic thesis that Fodor is building, is dependent on this ‘exhaustive’ relation, which precludes inferential role playing any part in semantic content.

4.2.4.1 Causal-cum-Nomological Relations and Semantic Access

A key part of what Fodor means by informational semantics is that “what bestows content on
mental representations is something about their causal-cum-nomological relations to the things that fall under them: for example, what bestows upon a mental representation the content *dog* is something about its tokenings being caused by dogs.”\(^{367}\) So it is in standing in a causal relationship with some property of the world, or whatever it is that a concept represents, such that there is a reliable tokening in the brain of the symbol associated with that concept, that makes that symbol refer to that thing. Information is about the *world*, and meaning can thus be re-described as information; “Meaning is information”\(^{368}\). This is what Fodor means by ‘informational semantics’.

Note that it is merely a matter of good engineering that there is a guarantee of lawful correlations between co-extensive things-in-the-world and mental tokenings corresponding to that thing-in-the-world in the head. Dogs cause DOG concepts because we’ve been designed that way - it’s not a logical relationship but a contingent one;

... how are certain lawful mind-world correlations (the ones that informational semantics says are content-constituting) achieved and sustained? ... [this is a question] about not metaphysics but engineering ... Answers to this engineering question can unquestion-beggingly appeal to the operation of semantic and intentional mechanisms, since ‘semantic’ and ‘intentional’ are presumed to be independently defined.\(^{369}\)

The world *just is* such that the properties that we ‘lock on’ to co-vary with tokens in our head; the causal relation between the two is an empirical fact about the world, it is a nomological necessity, not a conceptual one. What it is to acquire a concept is “getting *nomologically locked* to the property that the concept possesses”\(^{370}\)

The kinds of causal processes that provide semantic access are varied, and can even be constructed. Listening to gossip or experts is one way to get access to some mental token, as is looking through a telescope. There can be conceptual- or technical-mediation in semantic access; “Gossips, experts, witnesses, and of course, written records have it in common that each extends, beyond the sorts of limits that merely perceptual sensitivity imposes, the causal chains on which achieving and

\(^{367}\) Fodor, 1998, pg. 12  
\(^{368}\) Fodor, 1998, pg. 12  
\(^{369}\) Fodor, 1998, pg. 78  
\(^{370}\) Fodor, 1998, pg. 125
sustaining semantic access - hence conceptual content - depends”\textsuperscript{371} . I take it Fodor would endorse the strong claim that - having never seen a dog, or had any causal relations with dogs before - an expert saying the word “dog” to me would necessitate a mental tokening of DOG such that when I repeat the work ‘dog’ back to her, she and I are talking about the same thing. I refer to this example in \textit{Analysis}.

\textbf{4.2.4.2 Inductive Learning, Stereotypes and Innate Mechanisms}

Fodor’s causal-cum-nomological link with properties in the world is meant to enable informational semantics to avoid two related conceptual pitfalls. He does not want the acquisition of concepts to be inductive (on pain of circularity; “The problem with the theory that the primitive concepts are learned inductively was that it’s circular”\textsuperscript{372} ) and he does not want to necessitate a kind of massive nativism. If a concept is learned inductively - and one is committed to a “cognitivist account”\textsuperscript{373} of concept possession, whereby possessing a concept is a \textit{knowing that} rather than the non-cognitivist account which is a \textit{knowing how} - a problem of circularity arises in which one needs to know that something is in order to start the inductive chain. This would apply to whatever concepts one posits as the basic, or primitive, set (such as the Empiricists’ sensorium). This is a Sellarsian-flavoured argument; to take something as something at the start of the inductive chain, you already need to be able to identify what that something is - you are already required to take that something “under \textit{intentional} (rather than psychophysical) description”\textsuperscript{374} . So if you want an inductive account of concept acquisition, you will be committed to concept nativism w.r.t. the primitive concepts. Further, if you think concepts are atomic then the “special case [of the primitive] becomes alarmingly general”\textsuperscript{375} . One normally breaks the viscous circle by swallowing the fact that the primitive concept basis \textit{is} innate, but to maintain atomism (which Fodor’s RTM requires), one falls into a kind of massive nativism. Concepts that are atomic cannot be composed of more primitive, native ones, since they are – by definition – atomic.

Nor does it help one to avoid nativism if one makes a move to an account in which “having a concept is knowing how, not knowing that [and in which]... concept acquisition is arguably \textit{learning how} rather than \textit{learning that}, and [in which] it isn’t obvious that learning how needs to be

\textsuperscript{371} Fodor, 1998, pg. 78
\textsuperscript{372} Fodor, 1998, pg. 129
\textsuperscript{373} Fodor, 1998, pg. 124
\textsuperscript{374} Fodor, 1998, pg. 131
\textsuperscript{375} Fodor, 1998, pg. 123
inductive." Fodor dismisses this idea; “in lots of cases, it appears that how-learning itself depends on that-learning.” In addition, once one has attained the knowing how that is associated with some concept, “the particular skills that concept possession is usually supposed to implicate are perceptual and inferential, and these look to be just saturated with knowing that.”

Distinguishing knowing how and knowing that is, for Fodor, “a distinction without a difference.” Since I view concepts as a kind of skill, I am committed to such an account, and have tried to avoid the problem of circularity using the ‘similarity’ resources of Chapter 3. For more on this issue, please see Fodor, A Sympathetic Reading, below.

Hence, Fodor wants a “locking model of concept possession”, where the locking relation is more than merely “brute causal” and it results (as a matter of good engineering) in “typically having the right kinds of experiences”. The problem, he admits, in dropping the intentional character of the relation and allowing it to be just ‘brute causal’, there seems no reason why the same property (say ‘dogs’) always causes the DOG-token and not some other (say ‘giraffe’).

This, I take it, re-iterates the Sellarsian position: ‘brute causal’ is not enough to get the object under intentional description.

Most, if not all (it’s hard to read him here) properties that correspond to our concepts are ‘appearance properties’, like red. What it is to be a dog, or a doorknob (or, presumably, an instance of knowledge) is constituted by “how it strikes us” and “.. the point about appearance properties is that they don’t realise the questions that definitions, real and nominal, propose to answer: viz. ‘What is it that the things we take to be Xs have in common, over and above our taking them to be X’s?’”

For Fodor, this is a result that is derived from two constraining ideas; first, there does not seem to be a way of defining concepts, and so no way of agreeing in what way it is a complex object, and yet the same across all instances (I agree, see Chapter 2) and second, it does not appear to be

376 Fodor, 1998, pg. 124
378 Fodor, 1998, pg. 125
379 Fodor, 1998, pg. 125
380 Fodor, 1998, pg. 126
381 Fodor, 1998, pg. 128
382 Fodor, 1998, pg. 127
383 Fodor, 1998, pg. 136
384 Fodor, 1998, pg. 135
something simple, or primitive. If objects were complex “then [it] must have a definition” (which both he and I have argued it does not), and if it’s “un-definable, that must be because being an [OBJECT] is a primitive property. But, of course, that’s crazy ... Is it that being an [object] is ontologically ultimate? You’ve got to be kidding.”

The solution to this dilemma is to say that “[something like] an [object] is constituted by the kind of experience that leads to acquiring the concept [OBJECT]”. That being a thing “is a property that’s constituted by how things strike us, then the intrinsic connection between the content of [the thing] and the content of our [thing-experiences] is metaphysically necessary, hence not a fact that a cognitivist theory of concept acquisition is required ... to explain”. Here, Fodor wants to take apart the notion of something providing evidence for a concept, and something constituting a concept. Fodor avoids the circular cognitive trap (above) by avoiding the evidential relation between the object and the concept, which implies an inductive learning procedure:

I’m supposing, such an explanation is cognitivist only if it turns on the evidential relation between having the stereotypic [object] properties and being an [object] ... My story says that what [objects] have in common qua [objects] is being the kind of thing that our kinds of minds (do or would) lock to from experience with instances of the [object] stereotype

Fodor brings what he believes to be a statistical notion to bear - the stereotype - in trying to separate the object from objecthood. Remember, he wants his account to be non-cognitivist and hence he can’t use the concept OBJECT to lock onto the object itself:

... what I want to say is that [objecthood] is the property that one gets locked to when experience with typical doorknobs causes the locking and does so in virtue of the properties they have qua typical [objects] ... We have the kinds of minds that acquire the concept X from experiences whose intentional objects are properties

---

385 Fodor uses a ‘doorknob’ as his most-often-used example, presumably because - since one of the positions he is arguing against is nativity- a concept of a doorknob is patently not innate. If it works for doorknobs, it works for dogs and knowledge. I shall substitute ‘thing’ to make the argument appear less dependent on being an artifact - which it is not.
386 Fodor, 1998, pg. 135
387 Fodor, 1998, pg. 135
388 Fodor, 1998, pg. 134
389 Fodor, 1998, pg. 136
390 Fodor, 1998, pg. 137
belonging to the X-stereotype ... Stereotype is a statistical notion.\textsuperscript{391}

He believes that he can separate the notion of being an object from satisfying the object stereotype, that is, “the relation between instantiating the [object] stereotype and being an object is patently contingent.”\textsuperscript{392} The key idea is the following: a stereotypic object is something for which one can successfully enumerate the properties by which it is instantiated, for the objecthood itself, one cannot. Hence, the emphasis on the statistical notion of a stereotype. Learning by locking onto stereotypes is not a cognitivist story.

How far Fodor has moved away from the ‘prototypical’ account that he eschews is a little unclear - I take it that the move is to put all the composition and structure out in the world, with none in the head. He has moved to metaphysics the work that the cognitive scientist would have in epistemology, or cognition, or some kind of computation. Recall that prototypes don’t work for him because “it’s as certain as anything ever gets in cognitive science that prototypes don’t compose”\textsuperscript{393} and so - identifying compositionality as a desideratum - concepts can’t be prototypes. Here, the stereotype is much like a prototype, but it is something that one encounters in the world since objects can instantiate stereotypes- whereas prototypes are something in the head. All the compositionality of the features that make up the stereotype is a matter of metaphysics- the compositional work that defines what relations exist among features is in the world - that is one reason concepts can be atoms. However, it seems fair to ask: do stereotypes exist in the world in a way that prototypes do not, or has he manufactured them as a ‘just-so’ story - to enable him to separate the evidential from the constitutive?

The final point to this story is in defining just what it is that needs to be innate in order to be able to lock onto these infinite number of properties in the world. Fodor denies the content of the concept need be innate (it is not a cognitivist account, remember) but merely the mechanism by which one can lock onto the property; “the kind of nativism .... an informational atomist has to put up with is perhaps not one of concepts but one of mechanisms”\textsuperscript{394} (‘Just so!’ says the proponent of similarity-based neural-net accounts). Fodor likens the mechanism by which we can lock onto stereotypes to the sensorium; “the innateness of the sensorium isn’t the innateness of anything that has intentional content”\textsuperscript{395}. Fodor wants to separate content from content acquisition; “how much

\textsuperscript{391} Fodor, 1998, pp. 137-8
\textsuperscript{392} Fodor, 1998, pg. 138
\textsuperscript{393} Fodor, 1998, pg. 94
\textsuperscript{394} Fodor, 1998, pg. 142
\textsuperscript{395} Fodor, 1998, pg. 142
is innate in concept acquisition can be quite generally dissociated from the question whether any concepts are innate.” So, there may be lots of innate ‘stuff’ required to grab hold of these stereotypes, but the stuff that’s innate doesn’t have any intentional content. Sounds a lot like a job for the neural nets to me ... Again, see Fodor, A Sympathetic Reading.

Fodor is slicing the cognitivist - non-cognitivist distinction pretty thin, and I’m not sure it can be maintained. On the one hand, these mechanisms, the ones that enable us to lock onto stereotypes (which are what? things-in-the-world that we are capable of locking onto, but yet are not instantiated by any actual tokens of the thing-in-the-world?) sound a lot like skills to me. This would allow a non-cognitivist reading of concept acquisition, but at the expense of admitting concepts are not tokens in the head. That said, Fodor has made it very clear he does not want concepts to be skills, he wants concepts to be instantiated by little tokens in the head that represent content - presumably motivated by his psychological desiderata. So, on the other hand, with a very little ‘locking machinery’ one locks onto, precisely, the conceptual content that one needs to enter into the realm of publicity, systematicity, etc.

Fodor admits that his is a “what is called in the trade a ‘what-else’ argument” - it is the two constraining conditions above - the thing is not metaphysically simple nor is it complex - that led him to the metaphysically necessary relation between having a concept and being in a causal-cum-nomological relation to the property corresponding to that object.

Given the number of references to the Fodor, A Sympathetic Reading section in the last few pages, clearly I find something intuitively plausible about this kind of mind-world relationship, there is something about informational semantics that I am happy to adopt.

4.2.4.3 Differentiating Coextensive Concepts; It’s in the Head

One important addition to this causal-cum-nomological relation to things-in-the-world is the idea that “coreferential representations must be synonyms” Fodor is motivated to commit to this synonymy of co-referential representations by a need to deny any sort of Inferential Role Semantics (IRS). I outline three reasons why he wants to deny IRS in the next section 4.2.4.4 What’s so Bad About Inferential Role Semantics? First, however, I will explain why IRS is

---

396 Fodor, 1998, pg. 142
397 Fodor, 1998, pg. 133
398 Fodor, 1998, pg. 12
implied by non-synonymy of co-referential representations, and what is entailed for RTM by the
need for synonymous co-referential representations.

If “coextension is not sufficient for synonymy”\textsuperscript{399}, and one remains committed to informational
semantics, then there must be some “extra ingredient”\textsuperscript{400} that provides for the distinction between
the two concepts. The concepts would still co-vary (from the informational semantics commitment)
and the concepts are not synonymous (the starting premise) and so something must provide for the
distinction. A standard response is that a part of the concept identification is determined, at least in
part, by “what inferences one is prepared to draw or to accept”\textsuperscript{401}. \textsc{Conceptwater} is distinct
from \textsc{Concepth2o} because with one concept you are willing to infer that hydrogen is involved,
and with the other you may have no idea of the chemical makeup. Fodor will “call any theory that
says this sort of thing an Inferential Role Semantics (\textsc{irs})”\textsuperscript{402} \textsc{irs} must be abandoned for
independent reasons, given below.

So, the two concepts \textsc{water} and \textsc{h2o} are co-extensive, that is, they refer to one and the same set
of things-in-the-world, and for Fodor this means that the symbol in the head that co-varies with
each, will thus co-vary synchronously. He commits himself to idea that two concepts, call them
\textsc{Concepth2o} and \textsc{Conceptwater} have precisely the same content, they are synonyms in
his mentalese. \textsc{Conceptwater} means exactly what \textsc{Concepth2o} means. That is what it is
to be synonyms.

He will, however, commit that they are distinct concepts. Thus, distinct concepts can have the same
content; “I can’t, as we’ve just seen, afford to agree that the content of the concept \textsc{h2o} is different
from the content of the concept \textsc{water}. \textit{But I am entirely prepared to agree that they are
different concepts}”\textsuperscript{403} It seems odd, prima facie, since it is normally presumed that the content of
the concept \textsc{water} is distinct from the content of the concept \textsc{h2o}; in the Fregean tradition, for
example, they have different ‘senses’ but the same referent.

Fodor has now built the following picture: some single thing-in-the-world, that is a single thing by
virtue of having the same extension, has a causal-cum-nomological relation with a tokening in the

\textsuperscript{399} Fodor, 1998, pg. 13
\textsuperscript{400} Fodor, 1998, pg. 13
\textsuperscript{401} Fodor, 1998, pg. 13
\textsuperscript{402} Fodor, 1998, pg. 13
\textsuperscript{403} Fodor, 1998, pg. 15
brain of a single content, but which can be expressed by two distinct concepts. An example we have seen is the extension of water, which can be associated with two concepts with (necessarily) the same content: CONCEPT\textsc{water} (for a non-chemist who merely drinks and swims in the stuff) and CONCEPT\textsc{h2o} (for someone who is familiar with the table of elements and how it relates to the stuff she drinks and swims in). What is required is a method of distinguishing the concepts, without using any inferential analysis.

For Fodor what distinguishes concepts are Modes of Presentation (MOPs), which are (somewhat unhelpfully) almost anything as long as they are mental and they “distinguish distinct but coreferential concepts.” So there would be a MOP\textsc{h2o} and a MOP\textsc{water} whose role it is to present the concept to thought, and these two MOPs would be functionally distinct. MOPs are mental objects, functionally defined, such that it is first of all a method of presenting a concept to thought, and secondly a way of presenting that thought that successfully individuates that thought. One doesn’t think about MOPs, one uses them to think with; “it’s thinking with the MOP, not thinking about it”.

The identity of a MOP is constituted by what happens when you entertain it; it is functionally defined just to make distinct the concepts which have the same content; “what corresponds to the reasoner’s concept is not the mode of presentation per se [loosely, like a Fregean sense] but the mode of presentation together with how it is entertained.”

I take it that Fregean senses are meant to be abstract objects, the role of which is to distinguish different conceptual ways of referring to the same object. The standard example is the morning star and evening star -they are the same object, but are thought of in different ways. The problem for Fodor is that a sense is not sliced thin enough to individuate content. If there are two ways of referring to an object, say $a$ and $b$, then “the inference from ‘Fa’ to ‘Fb’ is never conceptually necessary.” The example Fodor gives is “Jackson is a painter” and “Pollock is a painter”. Thus Jackson and Pollock both have the same sense, that being a painter. But still, one may “wonder whether Jackson and Pollock were the same painter”. JACKSON and POLLOCK count as different MOPs, but they have the same sense. Thus, a MOP can’t be a sense. Fodor wants MOPs to be “sliced a good bit thinner than senses. Individuating MOPs is more like

\[404\] Fodor, 1998, pg. 15
\[405\] Fodor, 1998, pg. 18
\[406\] Fodor, 1998, pg. 18
\[407\] Fodor, 1998, pg. 17
individuating forms of words that it is like individuating meanings”\textsuperscript{409}.

Fodor goes on to commit himself to the idea that MOPs are themselves mental representations; “we might as well explicitly assume that MOPs are mental representations ... MOPs are among the proximal determinants of mental processes and mental processes are computations on structured mental representations”\textsuperscript{410}.

What Fodor will eventually commit himself to is indeed very language-like in the individuation of MOPs. I treat the issue of the language-like slicing of MOPs in the Conceptual Atomism section.

4.2.4.4 What’s so Bad About Inferential Role Semantics?

We have seen in the previous section that - by denying that inferential roles play a role in differentiating concepts - Fodor is lead to the synonymy of co-extensive concepts. The question to which I now turn is why, then, as part of his informational semantics, must Fodor deny a role to inference? Why must his informational semantics be mutually exclusive of inferential role semantics - in other words, why not a hybrid or sorts? He provides three reasons why not.

First, Fodor has committed himself to the mind being a Turing machine. Inference reduces to computation, “ie to operations on symbols”\textsuperscript{411}. As noted earlier, “For fear of circularity, [he] can’t both tell a computational story about what inference is and tell an inferential story about what content is.”\textsuperscript{412} Again, thought (inference) is defined as computation, and if computation implicitly requires a notion of symbol, then inference (computation) can’t fill in what that (symbolic) content is.

Second, Fodor again emphasizes the importance of nomic intentional explanation. For him, IRS has “holistic implications that are both unavoidable and intolerable”\textsuperscript{413} since “it’s pretty clear that psychological explanation can’t be subsumption under intentional laws if the metaphysics of

\textsuperscript{408} Fodor, 1998, pg. 16
\textsuperscript{409} Fodor, 1998, pg. 17
\textsuperscript{410} Fodor, 1998, pg. 22
\textsuperscript{411} Fodor, 1998, pg. 13
\textsuperscript{412} Fodor, 1998, pg. 13
\textsuperscript{413} Fodor, 1998, pg. 13
intentionality is holistic."**414**

Third, the main RTM argument is really just a warm-up for the main act which is “to argue for an atomistic theory of concepts”**415**. What this will amount to is that “satisfying the metaphysically necessary conditions for having one concept never requires satisfying the metaphysically necessary conditions for having any other concept.”**416** Here, it’s pretty clear that if the inferential role makes the difference between co-varying concept tokens, then Fodor can’t have conceptual atomism, since every concept must “*have* an inferential role”**417** and it takes more than one concept to draw an inference. I agree that IRS requires non-atomic primitive elements. So he can’t have atomic concepts, informational semantics and inferential role semantics. To save atomic concepts and informational role semantics, he has to drop IRS.

In doing so, however, he is really just getting his ducks in a row to attack a number of competing theories of concepts; “all three theories are really versions of one and the same idea about content. ... [they] all assume a metaphysical thesis ... namely, that primitive concepts and (hence) their possession conditions, are at least partly constituted by their inferential relations”**418**. If his account of RTM holds and is indeed the “only game in town”**419**, and it requires informational semantics, which in turn demands that concepts be atomic, then he has argued concepts must be atomic.

At the same time, however, there are at least two ways to save an IRS. First, I note the weakness is in his RTM - if it cannot be sustained, as I will argue it cannot, then the whole house of cards falls. No RTM, no need for atomic concepts. Second, he does admit that if there “is a decisive independent argument against the reduction of meaning to information”**420** then the need for an informational semantics goes. My approach, as I have pointed out a number of times, is to deny the larger, over-arching RTM thesis. I have laid the groundwork for such an argument in Chapters 2 and 3. The final bit appears in Analysis.

---

414 Fodor, 1998, pg. 13
415 Fodor, 1998, pg. 13
416 Fodor, 1998, pp. 13-14
417 Fodor, 1998, pg. 14
418 Fodor, 1998, pg. 35
419 Fodor, 1998, pg. 23
420 Fodor, 1998, pg. 14
4.2.5 RTM: Summary

A brief summary of Fodor’s position, as articulated thus far, would be helpful. RTM is the only game in town because it alone has the explanatory capacity to deal with a large number of explananda related our concepts. I have called these explananda psychological desiderata.

Two of these desiderata are the requirements of publicity and causality, which are in turn required by a nomic intentional explanatory framework. Fodor refuses to acknowledge other intentional explanatory frameworks, in which the mental realm is distinguished from the strict law-like behaviour of billiard balls and electrons by appeal to some other form of explanatory framework. Davidson’s anomalous monism is an example of an alternative form of intentional explanation, provided not as a positive argument, but to point out that it is unreasonable to deny that other forms of intentional explanation are tenable.

Three other desiderata are compositionality, systematicity and productivity. These are not presented as approximations of how we think, but are meant to reflect an empirical statement about, precisely, how we think.

I grant that RTM does indeed fully explain these desiderata as outlined here. A Turing machine, in which syntactically-defined symbols operate as concepts, over which computational processes implement thinking, and physical causation acts on the syntactic structure, thus preserving the truth of the symbolic content, would indeed produce nomic intentional behaviour and an accompanying law-like explanatory framework, and thought that is compositional, systematic and productive. I will not argue for this agreement, it seems clear enough. It’s hardly surprising, however, as RTM and the desiderata are really designed to fit together. Hence, my claim is that the over-riding strategy is one of conceptual coherence.

Semantic content is derived from causal-cum-nomological correlations with objects in the world, and the law-like correlations are a matter of good engineering (contingently existent correlations are not constitutive of the meaning of ‘semantic’ or ‘intentional’, as there are other logically possible ways of establishing those properties). This is what Fodor means by informational semantics. Inferential role semantics is not tenable for Fodor for three reasons; on pain of circularity, inference cannot play a role in defining content; IRS implies a kind of holism which would disallow nomic intentional explanation; atomic theories of concepts preclude the complexity of inferential roles playing a part in semantic structure.
What differentiates co-extensive concepts are MOPs, which are themselves mental representations. Computational processes on those MOPs constitute thought. MOPs are individuated (as we see below) along the lines of words in natural language. The atoms of thought are words.

4.3 Conceptual Atomism

RTM is meant to be the only game in town that can explicate the desiderata, and the only theory of concepts that fits the bill once one has adopted an RTM, is conceptual atomism. Fodor’s positive arguments for atomic concepts - aside from the fact that they are required by his definition of computation and informational semantics, above - is in denying various forms of inferential role semantics (IRS). Since my direct arguments against atomic concepts hinge on denying the RTM, I do not supply all of his argumentation against IRS, rather, I simply present Fodor’s theory of atomic concepts, and outline some of his main argumentative points. As I’ve noted before, if RTM falls, so too does his motivation for atomic concepts in the first place. My indirect arguments against atomic concepts comprise most of Chapters 2, 3 and the positive arguments for conceptual pragmatism. There are, as I’ve noted however, interesting points of agreement.

There are ontological as well as epistemic ramifications of conceptual atomism. By conceptual atomism, Fodor means that the underlying entities in the brain, those particulars whose machinations in the computational processes of the brain result in the language-like thoughts and the ascriptions of intentionality we make of each other, are atomic. The concepts which - in a very distinct way constitute our thinking and intentional states - are singular entities, and are literally tokened in the brain. By atomic, he means that there is no constituent structure - the concepts are basic; “... on the evidence available, it’s reasonable to suppose that such mental representations [of something like DOG] have no structure; it’s reasonable to suppose that they are atoms”\(^\text{421}\).

It turns out that Fodor will slice his mental atoms along the lines of the words in natural language. Remember that Fodor used MOPs (modes of presentation of concepts to thought) to distinguish between synonymous co-extensive concepts, like WATER and H2O. MOPs are themselves mental representations, and are individuated as atomic semantic items (conceptual content), which are in turn, individuated much like words in natural language.

\(^{421}\) Fodor, 1998, pg. 22
4.3.1 MOPs Slice as Words

One of the more glaring examples of how language-like Fodor would like his MOPs sliced, can be brought out by a treatment of what are called ‘polysemous’ words. Two examples of the same word in different sentences that have distinct meanings are ‘polysemous’; e.g. (from Jackendoff) ‘Harry kept the bird in the cage’ & ‘Sam kept the crowd happy’. Fodor denies that ‘keep’ has a common definition in the two sentences. Here, the word ‘keep’ expresses the same thing - namely *keep* - but the distinction in meaning between the two cases is “contributed by material in the underlying complement clause”\(^{422}\). It is the surrounding words that contribute to the MOP of keep, but keep always means *keep*;

In effect, what I’m selling is a *disquotational* lexicon ... and I can’t think of a better way to say what ‘keep’ means than to say that it means *keep* ... I know of no reason, empirical or *apriori*, to suppose that the expressive power of English can be captured in a language whose stock of morphologically primitive expressions is interestingly smaller than the lexicon of English.\(^{423}\)

In the above example in which ‘keep’ is meant to be polysemous - it is meant to have two different meanings - Fodor is simply denying that they *are* polysemous; “My theory is that there is no such thing as polysemy”\(^{424}\). There are reasons why such a word seems intuitively polysemous, but that is an illusion for Fodor. The reason that such a word *appears* (intuitively!) polysemous, is caused by “the assumption that there are definitions”\(^{425}\). What ‘keep’ means is *keep*, as simple as that.

What Fodor seems to be committing himself to is the notion that words that are not merely natural kinds (or categories that map easily onto an ontological extension - like water, or dog) but more complicated abstract words - extending the lesson from ‘keep’, to ‘knowledge’ ‘graceful’ or ‘tendentious’, for example - mean what they mean in a disquotationalist sense. ‘Knowledge’ simply means *knowledge*, and that concept is tokened by a MOP that is KNOWLEDGE - that’s all there is to it. Intentional content derives from the mental token itself and the understanding of that token. KNOWLEDGE is a concept that is atomic and the “metaphysically necessary conditions for having one concept never requires satisfying the metaphysically necessary

\(^{422}\) Fodor, 1998, pg. 54  
\(^{423}\) Fodor, 1998, pg. 55  
\(^{424}\) Fodor, 1998, pg. 53  
\(^{425}\) Fodor, 1998, pg. 53
Thus, what is tokened in the mind are semantic symbols that (pretty much exactly) follow the natural language. The distinction in meaning of the words across different sentences results from the contributions of the surrounding words - that is a part of the way the associated MOP is presented to thought, but “If, as I suppose, the concept KEEP is an atom, it’s hardly surprising that there’s no better way to say what ‘keep’ means than to say it means keep.”427 I am assuming here, and it seems a reasonable assumption, that I can substitute any of my favourite level-three pattern words for keep; ‘knowledge’ ‘ubiquitous’ ‘graceful’ etc.

4.3.2 Intuitions about Conceptual Connectedness and Analytic Truths

Fodor admits that there are very strong intuitions that certain words exhibit strong conceptual connectedness, most particularly words that would seem to admit of an analytic definition; “one the one hand, concepts can’t be definitions unless some sense can be made of intrinsic conceptual connection, analyticity, and the like ... on the other hand, there are lots of what would seem to be intuitions of conceptual connectedness, and that’s a prima face argument that perhaps there are intrinsic conceptual connections after all”428. For Fodor these intuitions are to be explained away; “I propose ... to try to explain the intuitions away.”429

The reason that Fodor believes we have these strong intuitions is that there are, indeed, connections between those concepts that appear conceptually linked, but those connections are of a metaphysical nature, not a conceptual one. Recall that the necessities of connection are metaphysical, which is one of the assertions upon which informational semantics is based; “What surely doesn’t embarrass informational semantics .. is the intuition that there is a necessary connection between ... being a bachelor and being unmarried ... informational semantics is a theory of content, and these necessities might all be viewed as metaphysical rather than semantic.”430. The mistake one makes in assuming that there are conceptual links is to mistake the metaphysical for the epistemic; “What you’re intuiting is really something epistemic.”431

426 Fodor, 1998, pp. 13-14
427 Fodor, 1998, pg. 55
428 Fodor, 1998, pg. 72
429 Fodor, 1998, pg. 72
430 Fodor, 1998, pg. 74
431 Fodor, 1998, pg. 86
The stronger one’s intuition that to get to some concept $C_1$ one needs to go through another concept $C_2$, so to get at BACHELOR you need to go through UNMARRIED and MAN, then the stronger one’s intuition that BACHELOR admits of an analytic definition, and the stronger one’s intuition that BACHELOR is not atomic, but has a complex constituent structure. Putnam has labelled such concepts “one-criterion concepts”, to get at the idea that there is really only one route the concept, that is - via other concepts. It is, for Putnam, ultimately a matter of convention that an analytic truth has a measure of the necessary; “Why is the exceptionless principle that provides the criterion governing a one-criterion concept analytic? ... they are true because they are accepted as true, and because this acceptance is quite arbitrary ... [there are no] systematic consequences beyond ... that of allowing us to use the pairs of expressions interchangeably.”

For Fodor, the appearance of the ‘exceptionless principle’, that which is grounded in the conventional for Putnam, is what explains the intuitions; it is an epistemic issue. That it is an epistemic, rather than a metaphysical or semantic issue, is certainly something upon which Fodor and Putnam agree. The way one gets semantic access to BACHELOR (or PAWN or TUESDAY) is normally through “inferences involving one or other member of quite a small family .. I’m suggesting that intuitions of conceptual connectedness are a sort of normal illusion; they depend on an understandable conflation between an epistemic property and a semantic one.”

Keeping our ultimate target of ‘knowledge’ in mind, I take it that Fodor’s position would be that ‘knowledge’ is atomic, and it means (of course) knowledge. That there are attempts to build and defend analytic definitions of ‘knowledge’ - the JTB analysis, for example - are reflections of strong intuitions that ‘knowledge’ is something like a one-criterion concept. However, the lack of consensus evidenced by the Gettier cases, may be interpreted as evidence that ‘knowledge’ is indeed atomic, and does not admit of definition.

---

432 Putnam, 1975, pp. 68-69
433 Fodor, 1998, pp. 85 - 86
4.4 Analysis

There are a number of ways to deny Fodor’s argument, and there are a number of somewhat surprising points of agreement. What I would like to do is argue against the Turing or computational aspect of RTM, but maintain a degree of sympathy for informational semantics and conceptual atomism. This is in reverse order of priority from Fodor’s analysis – which should not be surprising for his metaphysics is my epistemology, my epistemology is his metaphysics (it presently become clear what I mean by this statement). For now though, note that for Fodor, conceptual atomism follows necessarily once you adopt RTM – and you have to adopt RTM because it’s the ‘only game in town’ that can explain his psychologically-described desiderata. My strategy is to deny the desiderata themselves, thus eliminating the need for an RTM and the kind of conceptual atomism that follows.

However, sympathy for another form of conceptual atomism comes from psychology and the phenomenological character of real-world intuition. That view of psychology is what drives me to deny his desiderata in the first place. Note that my refutation of Fodor’s premise is what motivates me to adopt some aspects of his conclusion.

Fodor’s argument starts from the psychological desiderata as explananda in need of an explanation, and so if I can refute the desiderata on empirical grounds, then the raison d’etre of the RTM disappears. I have noted that it is a coherentist account (on concerns of theory structure), and it is sometimes unclear whether the desiderata are a result of the RTM, or if they are to be explained by the RTM. Either way, if they are refuted, so too is the RTM - there remains either no reason to explain them, or they cannot be the ‘output’ of the RTM. I first refute the compositionality, systematicity, and productivity requirements.

Fodor also consistently argues from the need for nomic intentional explanations to RTM, thus if I can refute the notion that intentional explanations are law-like in the strong sense that Fodor wants, or at least provide evidence that they need not be - shifting the burden of proof back to Fodor – then I have further eliminated the motivation for adopting an RTM in the first place.

In addition it’s claimed that the publicity requirement is needed so that communication is possible;

---

434 I say ‘surprising’ mainly because my view of the role of language, the type of cognitive processes that underlie thinking, and what sorts of entities comprise concepts (from an ontologic perspective) are all quite different - and all of these issues are each manifestations, for Fodor, of the central Turing thesis - which I am keen to deny almost
we need to share the same concepts in order to be talking about the same thing. There are many ways to fill the publicity requirement, and having the same mental token instantiated across persons is but one way. Instead of arguing at length for some other publicity conditions in detail, I instead argue that Fodor’s publicity requirement is too stringent, and that our communication is not quite as perfect as Fodor might think. We communicate, but success in so doing is a pragmatic affair, and the necessity of precisely shared references do not logically follow from such success.

I then turn to my sympathetic reading of a modified informational semantics and conceptual atomism.

**4.4.1 Denying the Compositionality, Systematicity & Productivity Desiderata; Trees and Minds**

Call these three desiderata, taken together - *structurality*. I certainly would not want to deny that thought has structurality; indeed, it is incumbent upon any theory of mind to explain *why and to what extent* thought has the structure that it does. What I deny, however, is that thought has the *kind* of structurality that Fodor insists it does. There is structure and there is structure.

Trees exhibit structurality, as do creatures, engines, water vortices and even clouds. Trees, for example, display compositionality - they are composed of atoms, molecules, or wood fibres, or fibres, branches, twigs and leaves, depending on your point of view. The syntax of a ‘tree’ is composed of sub-syntaxes of ‘fibres’, ‘twigs’, etc. in some loose sort of additive manner. Trees exhibit systematicity - the roots are always below the leaves, the leaves are predominantly on the end of the branches, the smaller branches extend outward from the trunk, the bark comprises an exterior shell around the nutrient-carrying inner wood, the water is carried to the top branches partly by the effects of capillary action and partly by negative pressure created by evaporation from the upper leaves, and on and on. Trees also exhibit productivity - I’ve seen some pretty big trees (displaying the same fundamental structure as little trees). This is not the structurality of a Turing machine, but that is the point – systematicity, compositionality and productivity are each a matter of degree and kind. The question is: to what degree do our minds exhibit these properties? This is, I take it, an entirely empirical question.

What Fodor is demanding is the *highest* degree of structurality - what I have previously called a ‘crystalline structure’ is what de Sousa calls ‘digitality’; “What Fodor [and Pylyshyn] are entirely.
actually demanding here, I suspect, is something else than mere systematicity. It is *digitality* [emphasis his]435 I follow De Sousa who describes digitality as the “substitution of a three-term relation for a two term relation”436, which is Platonic in form. What we do is “compare each thing to a paradigm and not to the latest in a series of copies ... That, I suggest, is what lies behind the insistence of Fodor ... and other defenders of the classical paradigm that there must be a language of thought437 that is systematic”438. What de Sousa has in mind is what you “really need the digital processes of *language* for is the cultural transmission of information, but not necessarily for that information that is processed internally”439. Just so - and this is a key point of the theory of concepts (pragmatic conceptualism) I develop in the next chapter – but the main point here is to characterize Fodor’s structurality as being of the highest degree. Fodor requires *digitality* to the extent that the smallest mental structures - which are themselves the ultimate constituents (semantic and syntactic) of extended intentional structures like conversations, long introspective machinations and books - are identical. The large-scale, public intentional structures inherit their content from unvarying small-scale physical mental tokens.

The question remains: is our thinking digital, or merely structured? This is an empirical question, and there is ample empirical evidence to indicate there is every reason to think it is not digital in nature. This should not be surprising - computers excel at precisely those tasks which we find difficult, and we find tasks facile that are notoriously difficult to program a computer to perform. To emphasize - I am not denying that we are capable of exhibiting some form of symbolic computation in thought (I, for one, can do quite complicated sums in my head), but that thought displays the kind of digitality that indicates that the digitality is basic, or constitutive of thought. To contrast, the digitality might be emergent, it might be visible only from a particular scale or view point or it might be an approximation only valid within some set of constraints (my own view is that it is an emergence of the second kind). So - what does the empirical evidence have to say?

If we were digital machines, to the bottommost causal rung of however it is that we describe that thinking machine, then we would display thinking characteristic of that sort of deep structure. We would, I take it, be perfectly rational, truth-preserving over large swaths of introspection, and so on. We manifestly are not. There are dozens of examples in the literature that show our biases and

435 De Sousa, 1991, pg. 6
436 De Sousa, 1991, pg. 6
437 I take for granted an established relationship between the term ‘language of thought’ and the Turing thesis presented here. If that is not clear, please refer to MacDonald and MacDonald, 1991 or Rand (2000).
438 De Sousa, 1991, pg. 6
439 De Sousa, 1991, pg. 6
reveal the unreliable nature of our thinking, and I only give one such example. This particular example is chosen because the two cognitive tasks are identical in logical form and it does not rely on biases that can be easily explained away (such as the ‘most recent evidence bias’ - a memory constraint? - or the kinds of bias exhibited in some moral choice examples - in which the different wording or presentation of the case could be indicate a different underlying conceptual structure upon which the Turing machine would operate). The example is meant to indicate a distinction between habitual, intuitive reasoning (pattern recognition) and logical algorithmic thinking.

The example is as follows, quoted at length from Dreyfus & Dreyfus:

Imagine that you have been asked to perform two tasks. In the first, you are presented with a stack of cards. One side of each card has either the letter A or the letter D. The opposite side of each card has either the number 4 or the number 7. The cards are now stacked with either side up, at random, and shuffled, so that thumbing through the deck you would see some A’s, some D’s, some 4s and some 7s. Your task is to determine whether or not the cards of this deck satisfy the rule “If the letter side of the card is an A, then the number side must be a 4.” To make that determination, you are to imagine that you are going through the deck, looking at the turned-up side of each card, one at a time, and turning over whichever cards you must, but only those cards, in order to verify or contradict the rule that every A must be accompanied by a 4.

... 

As task two, you are to imagine that you are the cashier at a supermarket and have the checks received that day stacked before you; some face up and some face down. Your supermarket has a rule. The checkout people are to accept checks for more than $50 only if approved on the back by the manager. Imagine that you are to go through the checks, one at a time, and turn over only those checks necessary to establish if the approval role has been followed.

... 

Why this pair of experiments? Because the two tasks are essentially identical. If you designate “over $50” as A, “not over $50” as D, “approved on back” as 4 and “unapproved” as 7, task two becomes task one.

The task with abstract symbols has a much poorer performance profile than the second, ‘familiar

---

440 Not our cultural transmissions, but our thinking.

160
context’ task. I do not believe this is an issue of limited resources - there is, if anything, less resources required on the first task as “not over $50” is conceptually more complicated than “A”. These results do not provide evidence that we are Turing machines, they provide evidence to the contrary. For further evidence of skill-based expert cognition, see Chapter 3. There are dozens of examples in the literature of our imperfect rationality, and so I will not press the point further.

For Fodor, it is a point of empirical fact that minds that are capable of thinking ‘Dave loves Mary’ have the ability to think ‘Mary loves Dave’. I don’t deny this fact - what I deny is that this “sort of symmetry of cognitive capacities is a ubiquitous feature of mental life” if by ‘ubiquitous’ it is meant ‘permeating the constituent structure of thought’ as opposed to ‘across all persons’. One can accept that the ability to think bFa implies the ability to think aFb, but that such ability is localized or emergent. By localized, I mean that the structure thus exhibited does not extend to larger and larger conceptual propositions or structures. By emergent, I mean that such structure is exhibited at one level - say of conscious thought - but does not necessarily extend down to whatever cognitive machinery is producing that thought. For Fodor, such symmetry permeates all thinking, down to the structures, upon which the computational processes that comprise thinking, operate.

Similar arguments can apply to productivity, although again the issue here can be pushed back to one of computational resources. We have quite small short-term memories, and thought may well be productive within that kind of band. But that is not an interesting claim, certainly not as interesting as the one Fodor wants to make - that “… there is no upper bound to how many times such operations [as thinking qua computation] may apply in the course of a construction”.

Alternative theories of mind certainly need to explain the limited productivity of thinking, but the ‘no upper bound’ idea - of limitless productivity - is one that we appear to have only by virtue of holding a pen to paper, or turning on our word processors. This reliance, indeed emphasis, on the need of external scaffolding in the form of such devices that extend our computational abilities, is embraced by pragmatic conceptualism. It is through the use of such devices, including being embedded in a socio-linguistic framework, from which our apparent digitality – and certainly our productivity - arises, not vice versa.

I take it that a clear example of cognitive productivity would be the ability to make sense of higher and higher orders of intentionality. Dennett’s ‘intentional stance’ is an example of first-order intentionality (the attribution of agency to an object in the world), and higher and higher orders are

---

441 Fodor, 1998, pg 26
442 Fodor, 1998, pg 27
not hard to describe: third-order – I knew that Chris thought that I believed X; fourth-order; I knew that Chris thought that I believed that Dawn felt Y; and so on. Dawkins points out in his recent book that there is a severe upper limit on our ability to entertain thoughts that attempt to establish very high orders of intentional structure in our thinking; “Very high orders of intentionality are probably confined to fiction”\(^4\), and he gives a wonderful example of a recursive application of intentional stances from the novel *Tin Men* by Michael Frayn; “Watching Nunopoulos, Rick knew that he was almost certain that Anna felt a passionate contempt for Fiddlingchild, and she knew too that Nina knew she knew about Nunopoulos’s knowledge ...”\(^4\). Here, with difficulty and only by staring hard at the page before me as an aid or tool, I am just about able to visualize what this sentence means – but I am severely strained in doing so (perhaps I only imagine that I can understand the sentence). This sentence reminds me of a game almost every child plays, saying ever-longer sentences to one another of the form “I know you know that I know that you know that he knows ...” until one inevitably breaks down giggling. What’s decidedly funny about the game is that you know perfectly well that at some point the sentence simply *makes no sense to you*, indeed, makes no sense to anyone (adults included). At some point, a clever friend breaks out a pen and paper to demonstrate that it does indeed make sense, but the need for pen and paper just emphasizes the point I’m trying to make. That clever friend is probably now a professor with a white-board on the wall, to which she makes constant reference.

The main point according to Fodor, though, is that thoughts exhibiting this kind of structure should not be difficult to understand as it is a natural extension of the systematicity and productivity of the processes that constitute the thinking process itself. I understand a response to this objection is to simply say: it’s a matter of limited short-term memory resources, and that, indeed, may be the case. However, there are two responses to this standard objection.

First, it doesn’t help to translate the intentional sentence above to a shorter syntactic item – relieving the load on short-term memory - such as “X’ing N, R y A z F, but A also y N y A N Y” (where y is a knowledge relation, z is feeling, etc.). The only way it may help is to reduce the amount of information on the page that you can *take in with one look*, which really just puts the emphasis back on external cognitive aids. So, it seems more fundamental than just a lack of short-term memory space. Secondly, Fodor is using productivity as empirical evidence that our thought displays deep structure, best *explained* by extrapolating that very same structure down to the constituent structure of thought itself. Productivity is supposed to be so rampant, so prevalent a characteristic, that the

\(^4\) Dawkins, 2006, pg. 183
\(^4\) Frayn, M., quoted by Dawkins, 2006, pg. 183
only way to explain it is to posit deep structure. The empirical evidence provided by our evident productivity is meant to be so incontrovertible that it can only be accounted for by way of causal processes that no neurologist has ever seen, and for which there is little other evidence. As such, the burden is on Fodor to explain why real world productivity is so limited, if it is a characteristic endemic of the thinking process itself. Limited memory resources may indeed limit the productivity of conscious thought, but until there is some further evidence as to the degree and way in which short-term memory plays that role, the evidence productivity provides in favour of deep structurality is mitigated by the limited degree to which it is displayed in conscious introspection.

That it is the world itself, and our real-time interactions with that world – rather than the existence of digitally perfect syntactic symbolic structures in the head - that provide the means by which we exhibit the structurality that we do, is not a new thesis. Dreyfus emphasizes that what is meaningful is not a map in the mind, but objects around us:

> When we are at home in the world, the meaningful objects embedded in their context of references [including intentional descriptive symbols] among which we live are not a model of the world stored in our mind or brain; they are the world itself. ...
> My personal plans and my memories are inscribed in the things around me just as are the public goals of men in general.\(^{445}\)

Haugeland wants to emphasize a kind of “intimacy of the mind’s embodiment and embeddedness in the world”\(^{446}\). For him, intelligence is not manipulating symbols in the head, but interacting with the objects, tools and semiotic elements around us (like the professor with the whiteboard, or the kid with the pen and paper):

> Until recently, however, [research into intelligent behaviour] has retained the assumption that the relevant ‘furniture of information’ is implemented as complex symbol structures that are, in many respects, just like the contents of the traditional Cartesian mind. In particular, they are internal to the individual agent ... If the significant complexity of intelligent behaviour depends intimately on the concrete details of the agent’s embodiment and worldly situation, then perhaps intelligence as such should be understood as characteristic, in the first instance, of some more comprehensive structure than an internal, disembodied “mind”, whether artificial or

\(^{445}\) Dreyfus, 1999, pp. 265-266

\(^{446}\) Haugeland, 1998, pg. 208
natural.447 [and] ... the most obvious externalization of human intelligence – texts,
images, maps, diagrams, programs, and the like – not because I underestimate their
importance, but because they are so similar to what is traditionally supposed to be
in the mind. ... As our ability to cope with the absent and covert, human intelligence
abides in the meaningful – which, far from being restricted to representations,
extends to the entire human world. Mind, therefore, is not incidentally but intimately
embodied and intimately embedded in its world.448
To emphasize the distinction: for Fodor, the syntactic mental items are primary, and meaning,
intelligence and - in particular- our exhibited mental structurality (which is, for Fodor, what I’ve
called digitality), are derivative of these items. The alternate position I am pointing to emphasizes a
different order of influence; the structure in the world around us, including both physical objects
(like fish and apples) and semiotic objects (like language), is primary and the structurality of our
thinking is derivative of these items. It need not be a one-way street, however, and there may
(indeed, probably is) an interesting dynamic of mutual influence: thought and ordered world,
ordered world and thought. For Fodor, though, there is only one direction of influence, and it is that
stark thesis and the kind of ‘thinking equals internal digital computation’ that I am denying.
It is the question of digitality and its relationship to mental activity that I am concerned with here. Is
digitality constitutive of human thought, or does human thought approximate digitality? What is the
nature of the structurality of our minds – is it more like trees or computers? This question is, in
turn, directly related to the question of computation and language - how is symbolic computation
and the use of language related to the underlying connectionist structure? I grant that there is such
computation, that there is symbolic-like thinking (and for present purposes let us assume
computational thinking is equivalent to language, thinking in bits like words). What I am denying
is that it is digital as such through and through. There are, as I see it, at least three possibilities:
a) symbolic computation constitutes thinking, as Fodor insists.
b) symbolic computation is an emergent activity, but it emerges as a perfect rendering of a Turing
machine.
c) symbolic computation is an emergent activity, and is best thought of as an approximation to a
Turing machine; that is, the emergent symbolic activity is roughly symbolic.

447
448

Haugeland, 1998, pg. 211
Haugeland, 1998, pp. 236-237
164


I think option c) is the most likely option (as I hope to have made clear in Chapter 3), and the causal structures involved in thinking are more like a tree than those of a digital computer. To be more precise, thinking is probably more like a vine, the wires, posts and constant pruning being analogous the structural effects of language. I do not claim to have proved this, only to have shown it to be a live option.

4.4.2 Refuting (the need for) Nomic Intentional Explanations

Much hinges on the need for law-like intentional explanation for Fodor. He consistently uses it as leverage to demand an RTM, for an RTM is the only “game in town” that can provide such an explanation. I will accept that RTM allows for nomic intentional explanation, indeed I’m even willing to concede that it may well be the only game in town that could provide such an explanation. What I deny is that there are not alternative intentional explanatory mechanisms that are not only made valid by their usefulness (predictive capacity, coherent relationship with other explanatory schemes, etc.) as an explanatory framework, but are probably more likely to be true, since RTM itself is already on pretty shaky ground.

There are two ways to approach this issue. One way is to deny that intentional explanations, as they stand, are actually nomic in character. Intentional explanations are not law-like scientific explanations, they simply don’t work the same way. A true intentional explanation is not law-like. A second way is to identify an alternative explanatory framework that has some intuitive plausibility, and thus shift the burden to Fodor to explain why such strict lawfulness is so essential a characteristic of proper explanation. I shall point to arguments that support each of these approaches, rather than give the full arguments myself, as these positions have been well established in the literature.

That intentional explanations are not law-like, or that there are competing or modified explanatory frameworks that do not follow the strict deductive-nomological form of scientific law, has been argued by any number of philosophers (Davidson, Horgan & Tienson, Place, Haugeland, etc.). What I want to do here is to motivate a reconsideration of what law-like might mean when it comes to intentional explanations, and soften the demands on what kinds of mental machinery may produce human behaviour so described. Intentional explanations are typically couched in belief-desire (B-D) explanations, and the question that I want to pose to Fodor is: what is the nature of such explanations, and why would we assume they are of the same form as scientific explanation to the extent that only an exception-less, precise Turing machine is capable of producing such
behaviour?

It is true that the question of the nature scientific explanation and its relation to covering laws is a complicated one, and while it could be argued that I do not do justice to this complicated issue, I will bypass much of the debate by simply asserting some contemporary views that capture enough of the relevancies of the subject to stand as – I claim – more than simply a straw man.

Horgan & Tienson (H&T) want to retain much of the classical symbolic framework that Fodor would like, such as “... structurally complex mental representations.... [and] mental representations have syntactic structure” but they would deny that the type of computing that the mind performs is of the kind that a Turing machine is capable of performing; “human cognitive transitions do not conform to tractably computable functions, and hence, cannot be attributed to the execution of programmable representation-level rules”. What H&T are pointing out is that “Human cognition is open-ended ... [there is] The potential relevance of anything to anything ... individual belief fixation [is] determined in part by ...factors that depend upon global properties of the current belief system ... [and] No one has any idea how cognition ...might be simulated using programmable representation-level rules”. Clearly H&T share a view I have developed throughout Chapters 2 and 3, and clearly they agree that a Turing machine is inadequate for the task of describing human cognition and behaviour.

The reason that this is so is quite simple. Intentional explanations are inherently and thoroughgoing ceteris paribus explanations; “We think that the important theoretical generalizations of intentional psychology are or should be ceteris paribus generalizations ... what we call ‘soft laws’”. These sorts of generalizations are of a distinct logical form, a form in which all exceptions to the general rule must come with an explanation. They are not “approximations to (or gestures towards) genuine laws” since the result of such a move would be “lost generality and failure to characterize fundamental tendencies of cognitive systems”. Their example is a ceteris paribus generalization of the form ‘Cheetahs are very fast runners’. This is a different kind of statement than either ‘Some cheetahs are very fast runners’ or ‘All cheetahs are very fast runners’. It is not falsified by the appearance of lame cheetahs, as the exception comes with a qualifying explanation. The number of potential exceptions and qualifications are limitless, yet the

---

449 Horgan & Tienson, 1999, pp. 10-11  
450 Horgan & Tienson, 1999, pg. 11  
451 Horgan & Tienson, 1999, pp. 11-12  
452 Horgan & Tienson, 1999, pg. 14  
453 Horgan & Tienson, 1999, pg. 15  

166
generalization loses no explanatory power. The conditions that sit in the background as ceteris paribus clauses, that allow the intentional explanation to go through, are essentially limitless in character. For H&T, “a true ceteris paribus generalization says something about the nature of the kind to which it refers, but since this is at the level of intuition, it does not carry any specific metaphysical commitments concerning ‘natures’”\(^\text{455}\) (I note with interest their use of the word ‘intuition’ here, but will not pursue it). Such explanations are, for H&T still good deductive nomological explanations since one often has “good, noncircular empirical reason to believe that ceteris is, was, or will be paribus”\(^\text{456}\). So here, we have a situation in which intentional explanation is still law-like, but incompatible with a Turing view of mind.

Whether or not the explanation remains nomological in form, however, is I think a matter of the degree and scope of the ceteris paribus clauses. If they are open-ended in number, and quite general in scope, then it seems to me that it’s the c-p clauses that do the real work, not the law-like character of the explanation. It is true that scientific laws also are ceteris paribus - Nancy Cartwright\(^\text{457}\) has argued vigorously that all of science is actually of this character, and thus scientific laws are not fundamental; “I believe we should read all nomologicals, as a ceteris paribus law .. they obtain only in very special circumstances: they obtain only when a nomological machine [that which defines and limits the ceteris paribus conditions] is at work”\(^\text{458}\) - but the point is that the ceteris paribus conditions are here finite in nature, and are not the dominant condition. Indeed, if the very model of nomologicality – scientific explanation – is not as pristine in its admission of only clear universally applicable laws and their accompanying theory, so much the worse for the aspirations of psychology. If there is no limit to the conditions one must place as background constraints, then it is in the defining conditions of those constraints that one is finding law, as much as in the explanation itself. Intentional explanations admit of infinite exceptions, and require unlimited ceteris paribus conditions to hold and it is the potential relevance of anything to anything that makes this so. I hope this point finds support in Chapters 2 and 3, and so I will not belabour the point here.

\(^{454}\) Horgan & Tienson, 1999, pg. 15
\(^{455}\) Horgan & Tienson, 1999, pg. 14
\(^{456}\) Horgan & Tienson, 1999, pg. 16
\(^{457}\) See also Cartwright, 1980 and 1999; “The impressive empirical successes of our best physical theories may argue for the truth of those theories, but not their universality. Indeed, the opposite is the case. ... Laws, where they do apply, hold only ceteris paribus. ... Our most wide-ranging scientific knowledge is not knowledge of laws but knowledge of the natures [tendencies of things to act in a particular way if the c.p. clauses hold] of things.”, 1999, pg. 4
\(^{458}\) Cartwright, 1999, pg. 25
Place, in responding to H&T, goes one step further in wanting to discard B-D explanations altogether, at least as far as being a discrete, isolable, and direct cause of the person’s behaviour. They are useful generalizations, and certainly help in understanding or predicting the behaviour of the human objects that surround us, but they are limited in their applicability to extrapolations to mental causation. Place has a sympathy for behaviourism “which condemned belief-desire explanations of behaviour as unscientific mentalism”\textsuperscript{459}. The problem with B-D explanations is “they characterise the dispositional determinants of behaviour on the assumption that there is a consistent and rational connection between what agents say about the situation confronting them when they have no motive to dissemble and what they subsequently do when confronted with a situation of that kind”\textsuperscript{460}. In other words, what we say about our behaviour and what causes our behaviour may not match. Indeed, this would appear to be the case particularly because “no one, to my knowledge, has ever even claimed to have recorded and translated a sentence in such a language [one that matches sentences to action], let alone had that claim substantiated”\textsuperscript{461}.

Behaviour, for Place (following Skinner), is contingency-shaped rather than rule-governed; “...contingency-shaped behaviour is behaviour that is shaped and moulded by past encounters with the actual contingency, in other words, by personal experience of the immediate consequences in one’s own case of doing one thing rather than another under the relevant antecedent conditions”\textsuperscript{462} whereas “Rule-governed behaviour ... is behaviour controlled by a verbal or other symbolic representation”\textsuperscript{463}. The point, for Place, is that it is an unwarranted assumption that the predicates of a natural language B-D explanation are actually instantiated in the brain and cause our behaviour. An alternative explanation that he provides is the contingency-based explanation, which is both motivated by and compatible with, a neural-network view of mind, devoid of substantive syntactic structures that match natural language.

It would seem that Fodor has opened the door to this type of view – open-ended computational resources required for context-dependent relevancy, massive ceteris paribus conditions required to bring out a useful generalization, substantive contingency-shaped behaviour based on past experience – in his admission that the perceptual conditions defining the identification of any particular natural kind are open-ended and infinite in character. He says, of being able to identify a dog: “If you try to list the sorts of perceptual environments in which dog-thoughts are likely to

\textsuperscript{459} Place, 1999, pg. 199
\textsuperscript{460} Place, 1999, pg. 200
\textsuperscript{461} Place, 1999, pg. 201
\textsuperscript{462} Place, 1995, pg. 70
\textsuperscript{463} Place, 1995, pg. 70
arise in a perceiver if he has the concept DOG at all, you will find that the list is, on the one hand, open-ended and, on the other hand, closely dependent on what the perceiver happens to know about, believe about, or want from, dogs.” However, Fodor simply pushes all of the open-ended kinds of conditions to metaphysics, and leaves the world to perform all that work for us (see my Sympathetic Reading for details). We just see dogs, and whatever complicated computations are required for that identification, are stripped out and made cognitively irrelevant in the tokening of a ‘dog’ symbol in our heads. Ultimately, it seems that a key distinction between Haugeland and Fodor is that for Haugeland, the complexity remains in the world (in which we are embedded) and for Fodor, that complexity is allocated (somehow) to our perceptual capacities.

What I hope to have accomplished here is to show that one of Fodor’s main claims – that the cognitive story we tell must accommodate a strong law-like intentional explanatory framework – is not as free of controversy as he would have it. I haven’t proven that intentional explanations are to be of a different kind, but I certainly hope to have cast doubt on Fodor’s characterization of intentional explanation.

4.4.3 Refuting (Mentally-Tokened) Publicity

Here, one can either deny that publicity is derived from mental tokens, or one can simply deny the publicity requirement as such, in the digital form as I have characterized the demand by Fodor. I will take the second route, although I will first note that there are plenty of options for the first.

The above (Section 4.4.1) references to the importance of external tools, semiotics and objects as enabling conditions (of a sort) is one flavour of getting the publicity requirement off the ground without requiring identical tokens in the head. Here, the objects by which meaning is made identical across speakers are objects that sit in the public domain, and meaning is obtained by reference to, agreement upon, and use of those objects. My denial of the publicity requirement relies on this general picture, and so I delay analysis until the next sub-section. What follows is merely a brief survey of some strategies, just to show that there are many routes to obtain the publicity requirement as demanded by Fodor.

Attempts to provide a publicity requirement that avoid commitment to mental-token identity across persons include various flavours of definitionism. Peacocke (1992) puts forward the idea that public definitions of concepts should be individuated by possession conditions. Possession

---

464 Fodor, 1998, pg. 79
conditions are “sets of inferences that a person must master to possess a concept” and those inferences may include “finding it compelling to make certain judgments under certain observational conditions.” Rey (1985) puts forward a similar account to Fodor, one that advances a version of informational semantics - that is, what makes dog concepts about dogs is the class of dogs - but wants to move the objects responsible for the publicity requirement out into the world and out of the head. What are in the head are ‘conceptions’, and they vary too much person to person. A definition for Rey, then, is a specification of whatever the metaphysical conditions of sufficiency and necessity are that specify the category in question. He has, thus, moved the informational requirement away from mental token identity.

Definitionism is notoriously plagued with problems (in this matter, I agree with Fodor). An immediate example is the fact that analytic philosophy is unable to actually come up with such definitions, the current problems with a justified-true-belief account of knowledge being only one such case. There is also an open question relating to the primitives; there must be some undefined primitives from which definitions are built. What are they - the disreputable sense-data that went out with the demise of logical positivism? Also - are they psychologically real? Does someone who talks of photons actually possess that concept in some form of the publicly-accessible definition? I also take note that I have built a picture of epistemology and metaphysics that is clearly positioned against a definitional view of concepts.

4.4.3.1 Denying the ‘Digital’ Publicity Requirement

I, however, want to deny the particularly strong publicity constraint put forward by Fodor - and likewise by anyone else who claims publicity must be of the degree Fodor demands. Like his structurality, the publicity he demands is a kind of digitality, a crystalline structure that admits of no roughness or ambivalence. Publicity, for him, requires that we share exactly the same concepts. I hope Chapters 2 and 3 have softened the reader up to be at least intuitively sympathetic to the following objection.

This spirit of this objection is markedly similar to that of the previous section. Just as intentional explanation will always require - to the point that they never really end - additional ceteris paribus conditions if they are to be law-like in their form, so too would the requirement that two people

---

465 Prinz, 2002, pg. 36
466 Prinz, 2002, pg. 36
mean *exactly the same thing*\(^{467}\) when talking, or communicating in some other way. Certainly *something* is shared, but need it be identical public concepts?

The argument runs something like the following. If two people are to communicate on Fodorian terms, they would have to have shared identical concepts. To communicate in real life is to be satisfied that you and your talking partner are ‘speaking of the same thing’\(^{468}\). If, to speak of the same thing, you had to have *identical concepts* (and be satisfied that you have, since the goal is communication), then you would never be satisfied that you had spoken of the ‘same thing’, and would have to talk forever to try and find that satisfaction. The reason one would have to talk forever to communicate on Fodorian terms is the same reason intentional explanations are in large part constituted by *ceteris paribus* clauses. One would have to add an infinite number of qualifications. We manifestly do not have to talk forever to nod in agreement, reaching satisfaction. We therefore communicate without identical concepts.

When I talk with my friend Chris about ‘statues’, for example, and endeavour to find out whether or not we are talking of exactly the same thing, when we deliberately put that conscious constraint on our communication, we do indeed find that our conversation does not end. Perhaps this is why dialogue within analytic philosophy, at least that part of analytic philosophy that embraces notions like necessary and sufficient conditions for a definition, does not end: it deliberately and self-consciously puts the Fodorian publicity constraint on dialogue. Philosophers work hard to ensure they speak of *exactly the same thing*, and find themselves always finding further qualifiers.

Fodor would deny this, of course, and the example I brought up earlier illustrates his way out. An expert says the word ‘dog to me. I have never had any previous semantic access to dogs - I’ve never seen one, heard one bark, read about one, etc. I trust the expert is an expert -she appears all grown up, has a serious look on her face and is wearing a smart pant-suit – I trust her. I repeat the word ‘dog’ back to her. Apparently, on Fodor’s view, we have just had a conversation in which real (the ‘hard’ publicity requirement is met) communication has taken place, substantiated by the fact

---

\(^{467}\) I have already argued (Section 4.2.1.a) that similarity is possible without identity (at least from an empirical perspective) and I do not rely on that particular argument here (although I note that it adds support to my position).

\(^{468}\) I take that this is what communication is in our world, and it is the only form of human communication with which I am familiar. Replace verbal communication with written, and nodding with some kind of assent, but the form remains the same: it is a procedure with epistemic boundaries, defined by those involved in the communication. Communication on the Fodorian-digital level may exist in some possible world, but not the one I and my fellow people inhabit. Unless, of course, the never-ending nature the dialogue in analytic philosophy is taken to be the paradigmatic case. Here, perhaps, due to the overtly introspective nature of the communication, the *goal* may be Fodorian, but whether that goal is reached is a larger question of which this thesis is but a part. Perhaps, for Fodor
that the expert and I have the same mental token in our heads. There is no need for the expert to explain what a dog is, nor point to one. This seems absurd to me.

I think Chapter 2 has made it clear that metaphysics is hard work - identifying what it is that we are talking about is not easy, it has to be earned. I deny that there are natural kinds, clearly delineated with crystal-clear metaphysical conditions of dog-ness, or tree-ness, and there are certainly not clearly-defined delineations of something more complicated, like graceful-ness. It seems to me Fodor has just helped himself to whatever it is he needs to make semantics what he wants. Even if my argument against natural kinds, and the like, does not go through in its entirety, surely I have made clear that one has to earn one’s metaphysical ticket. Simply asserting that there is a token, with identical semantic conditions across people, that refers to whatever it is we need to refer to, seems to me a free ride.

I think a perfectly reasonable view of communication is it can rest, ultimately, on agreement, and agreement rests on something like mutual nodding after a conversation. Each person agrees that agreement is reached, to the satisfaction of all involved, and that is what communication is at bottom - ‘close enough for jazz’, so to speak. This is obviously a very pragmatic view, and has an openly Wittgensteinian flavour.
4.5 A Sympathetic Reading of Atomism and Informational Semantics

Fodor and I part ways on a number of issues, as outlined in the previous sections, notably on the nature of intentional explanation, the nature of publicity and the role of language as well as the kind of structurality that underpins thought. Now though, I want to emphasize a more sympathetic reading of Fodor, while still maintaining those distinctions.

4.5.1 The Constitutive Nature of Experience

Take the following quotation from Fodor: “My story says that what doorknobs have in common qua doorknobs is being the kind of thing that our kinds of minds (do or would) lock to from experience with instances of the doorknob stereotype”\(^{469}\).

Take the following quotation from Chapter 3: “A pattern is defined as being anything which a pattern-recognizing creature would take as a pattern, and that pattern has thingness by virtue of that creature being embedded in a socio-linguistic community.”

There is clear agreement on the primacy of experience in defining, or conceptualizing, what an object is. There is a kind of irreducibility - we both take it as a constitutive analysis of the object that it simply appears to us as that object. There may be lots of ways to talk about the object, or study its parts, purpose or relation to other things - but what it is, at bottom, can only be defined in terms of a creature seeing it for what it is. We have agreed on a lack of analytic definitions for objects; “if [‘object’] can’t be defined, the reason that it can’t is plausibly not methodological but ontological”\(^{470}\). Objects are the objects they are because they appear to us as those sorts of objects.

In the case of something like a doorknob, this corresponds to the recognition of everyday objects as a form of O-Intuition; recall, from Chapter 1, “I am explicitly allowing the kind of judgments characteristic of everyday perception. Seeing an apple as an apple is such a judgment.” For more interesting patterns, well-developed H-Intuitions are responsible, but a ‘weak king side’ is just that because it is what the chess-expert sees as being a weak king side. There may be reasons that can be given in support, just as there can be reasons given why some particular object is a doorknob and

\(^{469}\) Fodor, 1998, pg. 137

\(^{470}\) Fodor, 1998, pg. 134
not a telephone. However, the defining conditions in some particular instance are not what makes it
an instantiation of weak king side, as that concept is more general and relies on global properties
that transcend any particular, situated, given instance – ‘weak king side’ is dependent on the other
player’s abilities, for example, which in turn depends on all kinds of historic, contingent encounters
with that player.

4.5.2 Mind-Dependency

Fodor has embraced the notion of mind-dependency, which is clearly one of my commitments in
pattern-recognition. What counts as a ‘pattern’ for me, depends to a large degree on the sort of
training one has encountered. There are two ways in which mind-dependency develops for me, one
defined by the scope of the global set of patterns (the number of ways in which the world has been
segregated into types of events) and one as formed by the path-dependency of the pattern-set
associated with one particular type. A musician hears different characteristics of the music by
having been trained to pick out the distinction between an A-minor and an A-diminished chord, and
a basketball enthusiast can ‘see’ a good pick-and-roll that a layman cannot. These are each
instances of the former, a result of the scope of the global set of patterns. However, the musician
could not hear that difference before she had already been trained to pick out the distinction
between major and minor. The basketball fan could not see the pick-and-roll until she had first
developed a sense of the rhythm of play. These are instances of the latter, the path-dependent
nature of mind-dependency.

For Fodor, mind-dependency comes in a number of forms. Firstly, it is the very fact that a mind has
locked onto the property that makes that property the content of a concept. Secondly, something
like ‘Tuesday’ is clearly mind-dependent, and is conventional; “So, no minds, no Tuesdays. But it
does not follow that there are no Tuesdays.” Further, “there are many properties that are un-
tendentiously mind-dependent though plausibly not conventional; ... being a convincing
argument.” I take it that ‘knowledge’ would fall under this category, and although he is loath to
admit that ‘knowledge’ is conventional in any way (since he presumably wants to retain a strong
publicity requirement across cultures), Fodor certainly appears sympathetic to the idea that what
makes something an instance of ‘knowledge’ is just as much the perceiving mind as it is the world.

471 Fodor, 1998, pg. 149
472 Fodor, 1998, pg. 149
Theoretical terms are not defined by the theory in which they are embedded, for Fodor, but “it’s just a concept that is, de facto, locked to a property via a theory [emphasis mine]”\textsuperscript{474}. It is by virtue of the mind holding a theory that it is capable of locking onto that property. How conventional a theoretical term is depends on how one views the nature of theories - taking a Quinian line, one could argue that the theoretical terms are conventional as the theoretical network writ large is itself conventional. My view would tend toward conventionalism, but Fodor’s does not, as he constantly emphasizes that all of the contingent relations among properties of an object are of a metaphysical nature. Mind-dependency and theory-ladenness thus come apart for Fodor. You need the theory to lock on to the property, but that the property is what it is, is an issue of metaphysics, not epistemology or theory-dependency. It is an open question to what degree Fodor thinks ‘knowledge’ is theory-laden, but it is certainly mind-dependent.

Fodor and I clearly part ways though, in that my view of mind-dependency can be seen as individual mind dependency for the same concept. I am perfectly happy that my concept of DOG is not the same as yours, and so I allow inter-personal differences in mind-dependency for the same concept. Fodor does not, motivated by his publicity requirement. I take it he gets rid of inter-personal differences by asserting that it is the stereotype that is locked on to, which is, presumably the same from person to person. For me, it is the actual training set that any individual has been exposed to, which is obviously contingent and person-dependent.

4.5.3 Ontology

Fodor: “Look, there is simply nothing wrong with, or ontologically second-rate about, being a property that things have in virtue of their reliable effects on our minds”\textsuperscript{475}. I am entirely comfortable with this statement, although Fodor is able to buttress his mind-dependent ontology with his view that the relations among properties (that make an object an object) is a metaphysical matter. It is a contingent matter, for him, that our minds pick out ontological categories by locking on to their properties via the effects those properties have on our minds. Those ontological categories exist nevertheless; Fodor is a self-admitted “Hairy Realist”\textsuperscript{476}. The mind-dependency I referred to above has no bearing on the ontological status of the objects themselves; “being mind-

\textsuperscript{473} They could not be so defined, as this would refute the atomistic nature of concepts.
\textsuperscript{474} Fodor, 1998, pg. 162
\textsuperscript{475} Fodor, 1998, pg. 148
\textsuperscript{476} Fodor, 1998, pg. 7, footnote.
dependent is perfectly compatible with being real.

I couldn’t agree more, if I could leave out all of the reliance on metaphysics. As I made abundantly clear in Chapters 2 and 3, the objects that inhabit our world – our ontology – is entirely dependent on our interests, our embodiment and our embeddedness. For me, the work is (for the most part) done on the epistemic side – our complicated pattern-recognizing neurology is what ensures that properties and objects are locked together into the relevant types – but the end result is the same. It is the phenomenological nature of O-Intuitions that provide for the furniture of the world, and that the bits and sticks of that furniture are what they are is irreducible beyond their appearance as such. Our epistemology defines our ontology; and so another way of presenting how these ideas embed within the quote from Fodor above would be: “being a property that things have [ontology] in virtue of their reliable effects [O-Intuition] on our minds [epistemic-dependency]”.

4.5.4 Stereotypes

For Fodor, it is by way of accessing a ‘stereotype’ that one obtains access to conceptual content. In my analysis of connectionist networks I shall eschew commitment to any particular brand of prototypes, exemplars or stereotypes in favour of a performance metric in which relevant similarities between objects is assumed (or given empirical grounding) - but it is shall become clear that a commitment of the connectionist position I adopt is that some kind of statistical averaging effect is at work. Fodor has the need to specify the “conditions for satisfying the [object] stereotype” because he is concerned with providing a means to separate the conceptual content from the means of content acquisition. I have no such concerns and so shall manage to remain neutral on the issue. That said, I am happy to adopt his ‘stereotype’ as that, somewhat abstract, epistemic object that our minds generalize to when recognizing a pattern

4.5.5 Semantic Access and Innateness

Here, I think, the most intriguing sympathies between the two positions can be found.

There are a number of instances where Fodor speaks of his “locking on” to properties in a way that is resonant with my “similarity” account of neural nets; what we get for ‘free’ as a creature

477 Fodor, 1998, pg. 147
478 Fodor, 1998, pg. 138
embedded in a life, in a world, equipped with neural nets is a similarity metric by which similar objects are *taken* as similar (without language) and *as an object* (with language). My position is that, given the empirically available evidence related to neural nets, it looks like what they give us is a similarity metric that, as a kind of computing primitive, locks onto characteristics of the world that are relevantly similar; that is, they lock onto relevant properties. My similarity metric is what provides for the “locking on”.

Fodor is loathe to “dump the load on Darwin”\(^{479}\) in which “[mother nature selects for the] type of relation between concepts and their experiential causes, .. by selecting a *mechanism that produces* that relation between one’s concepts and their causes”\(^{480}\). Fodor is stuck without this option because of the old problem of circularity - it seems you already need the concept to first identify the experience as one of the relevantly same kind. He has the problem of circularity because he cannot separate language from thought, he cannot separate language from the cognitive mechanism that does the work in building the relation between a concept and the causes of the concept. I can - the neural net provides for basic similarity as a complex kind of stimulus-response mechanism, and it is only by the introduction of language *on top of that pre-existing similarity* that introduces the notion of intentional content. Fish can tell food from non-food, but not *as* food and non-food. Language gets us that.

However, even respecting this distinction between us, Fodor’s talk is peppered with references to exactly the sort of mechanism I claim is instantiated by a neural-net, and he clearly wants that mechanism to be the only thing that is required to be innate. In other words, the only mechanism that Fodor will admit to being innate, and which provides for semantic access and concept acquisition, is functionally equivalent to my pattern-recognition stance toward the neural net. Fodor; “what has to be innately given to get us locked to [objecthood] is whatever mechanisms are required for [objects] to come to strike us as such ... the kind of nativism an informational atomist has to put up with is perhaps not one of concepts but of mechanisms.”\(^{481}\) “Too right!” says the pattern-rec enthusiast, “now what kind mechanism might that be?” Well, for Fodor, something like the sensorium; “the ‘innate sensorium’ model suggests that the question how much is innate in concept acquisition can be quite generally dissociated from the question whether any concepts are innate...The issue is whether it requires a lot of innate intentional stuff”\(^{482}\).

\(^{479}\) Fodor, 1998, pg. 128  
\(^{480}\) Fodor, 1998, pg. 129  
\(^{481}\) Fodor, 1998, pg. 142  
\(^{482}\) Fodor, 1998, pg. 129
It shall emerge that one important lesson the connectionists taught us, is that a neural net can lock onto a property. There are complicated issues that remain, like specifying the right training set and how language comes to play a role, and how to specify the relationship to conscious access, so forth (it’s not going to be *that* simple), but if we are not willing to admit into our conceptual theories the one, clear lesson from connectionism - that such networks can provide a non-intentional mechanism by which similarities can be constructed - then we have missed the point. The mechanism Fodor needs is a neural net. The reason he is loath to admit such disreputable company as connectionists into his intellectual party remains his commitment to the Turing model of mind. Some parties are better spoiled, and I read his ‘innate locking mechanism’ as an invite to the party. I turn now to developing the connectionist account in detail.

---

*Fodor, 1998, pp. 142-143*
Chapter Five

5. THE CONNECTIONIST STORY; IMPLEMENTATION MATTERS

5.1 Overview; C-Nets, Minds, Concepts and Intuition

Thus far I have argued that traditional computational devices – categorized as Turing machines that operate on the basis of symbol-manipulation – cannot perform the sorts of pattern-recognition that we humans do – and hence any view of mind that regards its operation as exhaustively described as a Turing device must therefore be false. Two lines of argument were given to this end; neither the metaphysics of patterns, nor the resultant epistemology of patterns, indicate that systems capable of exhibiting the pattern recognition behaviour of interest are amenable to implementation by such computational devices. Further, independent arguments were provided to doubt the psychological basis of such a view, contra Fodor. The obvious question that remains is, then: if not a Turing machine, then what? Connectionist networks (C-Nets) are here presented as a genuine alternative to Turing computation, and as such, are not susceptible to the same criticisms. Moreover, not only do such networks empirically demonstrate the capacity to recognize patterns of interest, but - once having bitten the bullet and taken this implementation story seriously - a number of psychologically suggestive corollaries follow with implications for how we may best treat the mind, not just as an intuitive pattern-recognizing system, but also as a vessel containing the mental co-relates of concepts.

First, I delineate the C-Nets of interest; they are non-linear, recurrent networks that make us of superpositional storage\(^483\), the models of which may also be extended to include – not just single networks – but a cascade of networks, or what I call a network-of-networks. The properties of C-Nets are then examined in a general, high-level manner in order to establish the claim that C-Nets are not merely Turing machines. Then, the lessons learned are extrapolated and applied to the questions about the nature of our own cognitive abilities; in particular, I articulate my account of pragmatic conceptualism – a theory of concepts – and make specific claims about intuition. I am not claiming that a Turing-type of computation is irrelevant to the mind; rather, the mind is not to be regarded as Turing computable through-and-through. Indeed, one of the explanatory obligations a

\(^{483}\) These terms shall be explained presently.
C-Net theorist must meet is to explain *how* the Turing-type of computations that are the cornerstone of our higher cognitive activities might emerge from the distinctive C-Net computing upon which such cognition is implemented. Such a burden is not fully met, but I shall point in some promising directions: language viewed as an internalized cognitive tool (rather than the product of inner systematicity), a more comprehensive network-of-networks model and a strong functional role assigned to consciousness.

My commitment to the *epistemic continuum* emerges here in the starkest of forms; many (but certainly not all) of the lessons and principles derived from C-Nets are gleaned from traditional (and by now hackneyed) examples that can best be described as *object* categorization, analogous to perceptual recognition. We are all by now familiar with the rock/mine detector⁴⁸⁴, and the features identified in Smolensky’s coffee cup example⁴⁸⁵. That said, a proper understanding of C-Nets extends to the role of labels (linguistic identifiers) *within* the computational economy of these networks, and I further extend that notion to the role such labels can play when implemented in a sort of cascade of networks, or network of networks. What I mean by these terms shall be made clear presently, but what the reader must keep in mind during the initial discussions is that perceptual-oriented examples are meant to be but the starting point of the general computing principles being discussed. Clark; “… we should insist that the connectionist is *not* committed to the use of only low-level perceptual feature to define the state space.”⁴⁸⁶ There is no reason to think *a priori* that the recognition of an object *qua* perceptual object is different in kind – from the perspective of C-Net processing - from some more conceptually-oriented example, the recognition of some situation *qua* intentional state, or some thought being judged an instance of justice, or knowledge.

C-Nets are well understood as a kind of statistical computer, generating clusters of similarly classified objects, but I here want to explore some deeper – and admittedly, more tendentious – interpretations of C-Nets. By taking it that ‘implementation matters’ my analysis *starts* from the premise – by now well-established by arguments about patterns themselves – that there are good *prima facie* reasons to think our cognitive abilities are underpinned by a C-Net-like structure. Now, I argue for repercussions generated from that premise: *given that our cognition *qua* pattern-recognition is underpinned by C-Nets, what else follows?* What is the role of traditional explanatory systems – like folk psychology – and what role in the mental economy do its posits play? What is a concept, both from the perspective of concepts as mental particulars and the

---

⁴⁸⁴ See Clark, 1993, pg. 41
⁴⁸⁵ See MacDonald & MacDonald, pp. 171-175.
perspective of concepts as publicly-shared linguistic items? What are the repercussions of this view for Intuition Driven Romanticism (IDR), which rests on the notion that it is deep within the heads of the folk that reside the publicly-shared conceptual content? Most central, what does this view have to say about the nature of intuition?

The mind emerges from this analysis as a very dynamic sort of thing, in which any ‘items’ in the brain that can be thought of as co-extensive with semantic content – the mental component of concepts – are best viewed as a kind of process, or as temporally fleeting activation signals, rather than as syntactic structures. The syntactic structure of the mind – in the form of a set of nodes and connections – is best viewed as a sort of accumulated knowledge or even as a theory of the world, but not in the sense that it is composed of some set of context-independent facts (or other informationally separable and distinct items, such a properties, features or prototypes). Rather, knowledge is seen as an overall capacity to react and generate relevant thoughts and facts.

As the mind undergoes the sort of development associated with learning, and generates a set of nodes, connections and weights, it gives rise to what is called strong representational change, a view on which the representational capacities of the mind change in ways unavailable to a Turing sort of device. Code and process are intertwined, and there is no sense in which computational processes can be differentiated from the code over which that process is meant to take place. Extending this anti-textual view of mind, a new, dynamic notion of explicitness is derived: that which is most explicit in the mind is that which is most easily activated, rather than that which is most readily accessible as text. The relationship of these claims to IDR should be clear (recall IDR is the claim that intuition is used to retrieve some deeply hidden but pre-existing piece of information about the term in play); there is simply nothing to find inside the mind, if what that means is to search around looking for some pre-existing piece of text as if one were searching in an attic for an old book. What one will find is the potential activation of thoughts, judgements or intuitions that are shaped by past experience (in the form of a set of connections) and by the current context in which the signal associated with that thought (judgement or intuition) is given form.

Here, my account of pragmatic conceptualism is given some flesh. There are two claims that I wish to accommodate. First, concepts are mental particulars that cause behaviour - causality. Second, concepts have content that is shared publicly - publicity. Concepts bifurcate along private and public lines on this view, and must be differentiated: concepts_public are associated with shared public and linguistically-identified content, and concepts_mental are associated with mental particulars.
Concepts are best thought of as a kind of counter-factual set of possible activation signals and are – as Heraclitus claimed about rivers – never the same twice. Concepts, however, are linguistic, publicly-accessible labels, whose content derives in large part from (some sort of) socio-linguistic behaviour and shared empirical evidence. These two aspects of concepts are related by way of causal-historical pathways with external stimuli, of which language is one part, and by a functional role provided by that greatest of mysteries, consciousness (or better, self-consciousness). It is by virtue of being aware of the linguistic label associated with the outcome of some activation signal (concept) that we are capable of attributing to it the publicly-generated semantic content. I admit that in bringing ‘awareness’ or ‘consciousness’ into play I’m passing off a large part of the explanatory burden to a somewhat murky place where many promissory notes are stored but few are ever paid off. Be that as it may, it is a criticism I accept by noting that one always seems to have to pay the piper sometime, and it is here where she plays for me.

Positive arguments about causality and publicity aside, any story about concepts that relies exclusively on the resources available to a C-Net theorist will not live up to the traditional demands of a comprehensive theory of concepts. It must be admitted that there remain significant limitations to C-Net-derived conceptual resources, particularly with regard to the question of how conceptual content is to be made generally available for the sorts of computation so well addressed by traditional LOT theories – the systematics and productivity requirements. Filling in that account is beyond the present purview, and I accept that C-Net resources need buttressing in ways I cannot provide (although I shall give some arm-waving accounts of possible routes to explore and I have already provided reasons to think those characteristics, presented as demands to the conceptual theories, are too strong – see Section 4.4.1). Intuition is of primary concern here, and pragmatic conceptualism is developed to the point at which intuitive judgements may be characterized within that larger framework.

An intuitive judgment is best viewed as an occurrent activation signal. Not all activation signals are intuitions, however, as the previously defined constraints of accessibility to consciousness and inaccessibility of the underlying inferential process must be met. Intuition bifurcates along the same lines as do concepts, and consciousness is assigned the same functional role: intuition is the linguistic form that the intuitive judgment takes, and intuition is the occurrent signal itself. Intuition is thus revealed to be an expression of one’s accumulated knowledge, in the form of a response to some situation. There is common ground among people in their knowledge - given our commonly shared world, interests and capacities - but there are also individual (and cultural) paths over which our individual minds (and cultures) have evolved, given a unique causal-historical path.
of encounters with the world and the accompanying unique set of connections. Intuition, as one element of the counterfactually defined set of signals that comprises a concept, can be regarded as evidence of one’s own concept. Intuition can be regarded, from a phenomenological perspective, as an awareness of the linguistic label generated as the transformed activation signal that is - to the C-Net - an *appropriate response* to some situation.
5.2 The C-Nets of Interest

I take it that the standard model of a series of interconnected nodes, with a set of input and output nodes along with a varying number of hidden layers, is familiar to the reader\textsuperscript{487}. While it’s true that C-Nets can implement Turing machines – indeed, simple linear networks cannot be anything but a Turing machine of sorts\textsuperscript{488} – these are obviously not the C-Nets of interest here. The general picture upon which I shall rely is not one of simple, linear input-output networks, but networks better described as recurrent, dynamic, nonlinear, superpositional and distributed (what I mean by these words shall be made clear presently). These C-Nets share a general form in which the output layer is connected to the input layer by way of one or more hidden layers, which may themselves be connected to an “additional set of context units connected to the hidden-unit layer”\textsuperscript{489} (to store the previous state of the hidden layer(s) - these are known as ‘recurrent’ or ‘bi-directional’ networks). The connections between nodes are not simply additive, but reflect a more complicated nonlinear relationship. Finally, the output layer encodes for a semantic interpretation of the input. Representation in such a network is to be thought of as the pattern of activity that passes through the network, rather than as some set of nodes or weights. This is an important move, and why it is justified shall be made clear presently (see Section 5.3.2).

These C-Nets of interest are a “certain breed of connectionist model” – distinct from those that may be interpreted as symbolic processors - and are “… distinguished by three features … superpositional storage, intrinsic context sensitivity, and strong representational change”\textsuperscript{490} and

\footnotesize
\textsuperscript{487} A brief description of a standard C-Net would be as follows – I shall be very brief as I do not want to detract from the general discussion by getting mired in the details of C-Nets that have been discussed at great length in the literature. An input is applied to the input layer, that signal passes through any number of connections to the hidden layer(s) which act to “preprocess the information” (Bechtel & Abrahamsen, 1991), and from there on to the output layer, at which point it can be interpreted. A typical network takes data in the form of a digitized signal, processes (or transforms) it through a number of layers, and the output is interpreted as a semantic assessment of the input signal (it was a ‘tank’ / ‘rock’, or in the famous NETTalk network a number of phonemes are represented at the output layer as a translation of text in a phonetic form). The equation governing the signal is the “law governing the interaction of the units in the network” (Clark, 1993, pg. 61) and is normally described by a dynamical equation (a differential equation) which describes the signal as it changes over time as it moves through the network. That governing equation is defined by the set of weights that describe the connections of the system, and the connections themselves. Such systems are taught to respond to relevant inputs in appropriate ways, by exposing them to a number of test cases, and ‘tuning’ the weights by way of a number of different learning algorithms, one of which is ‘backward propagation’ – in which one alters the weights of individual nodes by “propagating the error measure that is calculated at the output units back through the network” (Bechtel & Abrahamsen, 1991, pg. 86).

\textsuperscript{488} See Section 5.2.2

\textsuperscript{489} Clark, 1993, pg. 26

\textsuperscript{490} Clark, 1993, pg. 17

184
which are governed by a “nonlinear evolution equation”\textsuperscript{491}. Although each of these features play a role in differentiating C-Nets from their classical counterparts, it is “their use of superpositional representations [that are the] source of their genuine divergence from classical approaches and the source of the features … which make the models psychologically suggestive”\textsuperscript{492} I focus, then, on what is meant by superpositional representations (otherwise known as distributed representations), followed by each of the other two key features, and why C-Nets are to be regarded as distinct in kind from classical computing architectures. First, though, I note some other key distinctive marks of - and assumptions about - the C-Nets of interest, and in particular, I want to emphasize that any psychologically realistic view of C-Nets must move away from the intuitive limits of the toy examples often provided to get at the fundamental operations of C-Nets, and extrapolate those lessons to the vast and (surely) multiply-connected network of networks that is our brain.

5.2.1 The Role of Theory or, Not Just Perception

Discussion of C-Nets is normally facilitated by imagining that the input is a digitized signal of some perceptual event (a photograph, for example, of a particular, situated rose, dog, tree or the number ‘3’) and the output is a semantic interpretation of that signal in the form of a categorization function (the signal encodes the label of ‘rose’, ‘dog’, ‘tree’ or the ‘number 3’). Hence, the signal is applied to the input layer as a digitized image of a token of some item, it may pass through any number of layers and connections – including connections that loop back to previously encountered hidden layers – before hitting the output layer as an interpretable signal or a coded semantic label. Alternatively, these sorts of C-Nets can be run in reverse, in which an encoded semantic label is applied to the output layer, and the central statistical tendency of previously encountered tokens of this type can be generated. This is certainly a useful way to think of C-Nets, but it “limits our imagination to the perceptual, and limits our purview to that of a single network constituting a single semantic metric”\textsuperscript{493}.

The role of theory is somewhat contentious in C-Net analysis, but I shall make a case that there are two ways in which theoretical terms may play a role in C-Net computing. First, theoretical terms may be provided a role in a kind of cascade of networks, in which the output of one may be the input of another. That output is normally considered to have semantic, interpreted content, and when used as the input to another network, it can fill the role of abstracted, structured information about

\textsuperscript{491} Clark, 1993, pg. 61
\textsuperscript{492} Clark, 1993, pg. 17
the prior processed signal, and hence imbue some theoretical import. Second, there is a fairly straightforward way in which the set of connections and weights may constitute a theory of the domain to which it is applied\textsuperscript{494}.

To think that C-Nets cannot take into account theoretical considerations in making classification judgments is thus argued to be mistaken, and “the problem of theoreticity is a pseudo-problem”\textsuperscript{495}. How networks learn to pick out relevant theoretical or structural properties “out of what must in the first instance be perceptual information remains, however, … tricky”\textsuperscript{496}, but I take it for granted that the sort of general explanation provided below is satisfactory for the task at hand. Note that the ability to learn novel ways of interpreting data is an instance of unsupervised learning\textsuperscript{497}.

5.2.2 The Importance of Nonlinearity

Nonlinearity of the governing equation is important in two ways; first, as it relates to superpositional storage, and second as it relates to system performance. In terms of the former, Smolensky’s (1986, pg. 411-413) mathematical treatments of C-Nets shows that in any network governed by a linear equation – which really means that all of the weights associated with the nodal connections are simple, linear additive functions – “the distributed representations could be replaced by local ones without making any difference to the normal behaviour of the system. … In nonlinear systems (ie ones in which the evolution equation is more complex), this is not the case.”\textsuperscript{498} An example of a nonlinearity in a network would be a “squashing function at the hidden-unit layer”\textsuperscript{499}, the output does not have a single multiplicative or additive value over the entire input range, but it varies on the input’s value.

If a system is linear, then, any superpositional representations could be eliminated in favour of local representations without degrading the performance of the network. This is really just another way of saying that the network could be replaced by a Turing machine – it would not, then, be exhibiting pattern-recognition behaviour that could not, in principle or in practice, be implemented on classical

\textsuperscript{493} See Section 5.3.1
\textsuperscript{494} See Section 5.4.2.2
\textsuperscript{495} Clark, 1993, pg. 103
\textsuperscript{496} Clark, 1993, pg. 103
\textsuperscript{497} See Section 5.2.3
\textsuperscript{498} Clark, 1993, pg. 61
\textsuperscript{499} Clark, 1993, pg. 103
computing device. Thus, in order to qualify for the status of non-Turing computing, a network must have some degree of nonlinearity in its governing equation.

In terms of performance, another important feature of nonlinear systems is that “nonlinear systems can provably encode any possible categorization function”\textsuperscript{500}. What this means is that nonlinear systems can perform \textit{arbitrarily complex transformations of input signals} - there are no theoretical limits to the degree of sensitivity to any particular feature of the input - and so a nonlinear network allows for category judgments of enormous subtlety. C-Nets are often described in mathematical terms by referring to the topology of a state-space\textsuperscript{501}, and these mathematical models can be used to describe in semantic terms the performance profile of the network. Regions in state-space correspond to categories available for judgment, and what nonlinearity introduces is the means to eliminate “irrelevant and misleading correlations among the training inputs”\textsuperscript{502} by enabling the network to shape its response profile (pushing and pulling the topology of its state-space) in response to arbitrarily subtle similarities in the input cases.

It would come as no surprise to anyone who takes the parallels to the real neural network of the brain seriously that the governing equation need be nonlinear. If the C-Net is to approximate the activity of a real brain, then it would need to be approximating a very non-linear and analog chemical soup of neurotransmitters, inhibitors and transmission pathways.

\textbf{5.2.3 Supervised vs. Unsupervised Learning}\textsuperscript{503}

One key differentiator among C-Nets is whether or not they are trained up on a given set of data, with predetermined classes of objects, or whether they form independent decisions about the most relevant ways to classify the data. \textit{Supervised} pattern recognition is generally governed by the provision of a training set to the network. Ripley, a mathematician concerned with the nitty-gritty engineering problems associated with building useful C-Nets, defines the training set of a supervised network as “a collection of labeled examples”\textsuperscript{503}, and the job of the network is to “distil the essence of the grouping”\textsuperscript{504}. A much more versatile approach – and still very difficult to engineer – is a C-Net that is capable of forming novel categories or otherwise generating new ways

\begin{footnotesize}
\begin{itemize}
    \item \textsuperscript{500} Clark, 1993, pg. 103
    \item \textsuperscript{501} I shall have more to say about the role of mathematical treatments below. For an extended treatment of Smolensky’s vector analysis see Rand (2000). See also Section 5.4.2 below.
    \item \textsuperscript{502} Clark, 1993, pg. 103
    \item \textsuperscript{503} Ripley, 2005, pg. 3
\end{itemize}
\end{footnotesize}
of extracting structure from the world with which it is in contact. Ripley; “The discovery of new groupings is called unsupervised pattern recognition.” Here, Dreyfus’ accusation that AI researchers constantly employ ‘cheats’ in the form of human decisions as to relevance, problem type, etc. resurfaces. Supervised training assumes some pre-determined classification of objects into relevant types, and that classification imports human cognitive skills; the C-Nets are then really just acting as fast-track statistical computers, which can be quite easily (and fully) modeled in fairly straight-forward mathematical ways, such as cluster analysis. The difficulty in engineering unsupervised C-Nets can be seen in that most texts devoted to neural network engineering are almost entirely devoted to methods of building supervised networks, with the attendant problems of architecture/topology design choices, feature extraction, etc.

Difficult as it may be to build unsupervised networks, I take it that these are practical difficulties facing a new science and that the theoretical possibilities remain open, and philosophical implications of the practical difficulties are of little concern. Indeed, unsupervised C-Nets can be built, but it is very difficult to generate networks of any practical engineering use except within limited domains such as to generate confirmation of previous classificatory schemes. Ripley; “Unsupervised methods are used when no classes are defined a priori, or when they are but the data to be used to confirm that these are suitable classes.” It is quite possible to build any number of network configurations that take on the task of defining their own categories on a set of input data, and such networks are “generally designed for visualization, either to show views of the data which indicate groups, or to show affinities between the examples by displaying similar examples close together. … Groupings found by unsupervised methods are usually referred to as clusters.” In other words, unsupervised learning is not theoretically problematic, but the usefulness of any particular network’s interpretation of the data is dependent on its usefulness to us: did the network confirm that a classification we made is relevant from some other perspective? One example Ripley provides is the classification of a species of crab (Leptograpsus); “There the [human] division into species was based on colour, and the interesting question is whether this is supported by morphological differences [which the C-Net was trained to pick out].

Unsupervised learning plays a key role in the development and deployment of theoretical terms, and although – according to Clark - it “is perhaps a little harder to see … how a network might discover

504 Ripley, 2005, pg. 3
505 Ripley, 2005, pg. 3
506 Ripley, 2005, pg. 287
507 Ripley, 2005, pg. 287
508 Ripley, 2005, pg. 288
such a normative feature for itself – that is, how it might be forced to generate novel and highly theoretical representations at the hidden unit level … it is not obvious that generating [the theoretical term that plays a normative role] is a *qualitatively* different kind of task from [another, more standard] classifying principle.”

Here, Clark is attempting to underplay the theoretical problem associated with a kind of unsupervised learning, while admitting the practical problems in current C-Net modeling. I take this position to heart; while problematic from an engineering point of view, such learning is not a philosophically deep problem.

I shall assume then, that some sort of unsupervised learning is entirely possible and indeed, probably quite rampant in human cognitive development. Note, however, that ‘unsupervised’ learning, without qualification, is a much stronger notion that I require. A completely unsupervised learning process would not be constrained by, or answerable to, *any* outside process. Such a network would simply settle on *some* form of structure, or statistical invariance, that it happened to come across given it’s initial architecture and weights. A human, however, is constrained both by external and internal forces. External constraints would come in the form of parents and other humans who help us form useful categories, as well as the constraints of having to interact successfully with a world that has a particular structure for us (apples are good to eat, and stones are not). Internal constraints would come in the form of our particular perceptive systems, which presumably pre-shapes the signals hitting the various parts of our brain.

---

509 Clark, 1993, pg. 102
5.3 Why C-Nets are Not Turing Machines

I turn now to outlining the characteristics of C-Nets that differentiate them from Turing machines. I start by analyzing the source of C-Nets’ interesting behaviour – superpositional storage – in terms of what is called a ‘semantic metric’. A semantic metric is a relationship that holds between the vehicles of representation and the items that may be represented. From the semantic metric comes the notion of ‘subsymbolic’ computation, and that notion drives a discussion of the context-sensitivity of subsymbolic components. It will be argued that there are no context-free computational elements in C-Net computing, and hence C-Nets differ markedly from their Turing counter-parts. It shall also be argued that C-Nets differ in that code and computational processes are deeply interpenetrated, and the result is that the representational resources of a C-Net change in ways that the resources of a Turing machine cannot. This difference is marked by the terminology ‘strong representational change’.

5.3.1 Superpositional Storage and the Semantic Metric

Superpositional storage is a fairly straightforward notion, but the implications for system performance are far-reaching. A system is said to be fully superpositional if “the resources used to represent item 1 are coextensive with those used to represent item 2”⁵¹⁰ – and, *given the idea that it is the patterns of activity generated by the signal passing through the nodes of a network (governed by weights and connections) that are said to ‘represent’ the item which is being semantically interpreted* – then C-Nets are said to be superpositional to the extent that a “single set of weights … do multiple duty”⁵¹¹. A set number of weights and connections are modified (as the system is trained) so as to be able to produce a number of desired outputs (categorization functions) over some range of inputs. Thus, the very same set of weights and connections can be said, in a very literal sense, to be used as representational resources – as the representing signal passes through the network - for multiple semantic items. As new items are added to the networks capacity – that is, as it is trained to generate unique and relevant output signals for a number of new inputs – the weights are modified “in a way which preserves the functionality (some desired input-output pattern) required to represent [the existing item] while simultaneously exhibiting the functionality to represent [the new item]”⁵¹². The key move here is not merely that the

⁵¹⁰ Clark, 1993, pg. 17
⁵¹¹ Clark, 1993, pg. 18
⁵¹² Clark, 1993, pg. 17
representation is distributed or extended over many nodes\textsuperscript{513}, but the use of “internally structured extended representations”\textsuperscript{514}.

An ‘internally structured extended representation’ is one that “is in some sense a nonarbitrary construction … [and which] exploit(s) the extendedness in [a] semantically significant way”\textsuperscript{515} such that the system can be regarded as “imposing a semantic metric”\textsuperscript{516} on the inputs. The idea here is that the individual nodes encode for individual features common among the range of inputs, and the semantic metric captures the idea that “semantic … similarity between representational contents is echoed as a similarity between representational vehicles”\textsuperscript{517}. The metric is imposed on the inputs, however, because a typical network does not have individual units and the representational resources set in advance, rather they are ‘learned’ by the system and are often very difficult – if not impossible – to semantically interpret in a post hoc analysis. It is the network that decides, in an important sense, how to allocate the internal processing resources to best effect, and so how to ‘divvy up’ the semantic space into relevant individual features on which the inputs overlap in some way.

The example Clark provides to differentiate between local, distributed and superpositional storage is one in which the C-Net is used to recognize text, and I include this example for clarity. A local network could represent a letter by having a “a single unit stand for that letter … this is clearly a localist representational scheme”\textsuperscript{518}. Another scheme could be one in which “the letters are represented as patterns of activity across [many] units”\textsuperscript{519} and the joint activity of sets of units (say 1, 2 and 3 for ‘A’, 4, 5, and 6 for ‘B’ and so on) is used to encode individual letters. These are extended representations (several units are used to encode single semantic items) but “the scheme is still effectively localist, because the representations, through spread out, do not exploit that extendedness in any semantically significant way”\textsuperscript{520}. Finally, individual units may …

\ldots stand for features of letterforms in a given font, such as l, - and _. The system’s representation of the letter A can then be just the joint activity of the various features which distinguish it, and likewise, for B, C. etc. Here, at last, we are dealing with distributed representation in an interesting sense. Notice that in such a scheme the

\textsuperscript{513}\ Compare to simple, linear networks in which each node can be said to represent some item.
\textsuperscript{514}\ Clark, 1993, pg. 19
\textsuperscript{515}\ Clark, 1993, pg. 19
\textsuperscript{516}\ Clark, 1993, pg. 19
\textsuperscript{517}\ Clark, 1993, pg. 19
\textsuperscript{518}\ Clark, 1993, pg. 19
\textsuperscript{519}\ Clark, 1993, pg. 19
\textsuperscript{520}\ Clark, 1993, pg. 19

fact that the letterform E shares more features with F than it does with C will be reflected in the system’s use of resources to code for the letters.\textsuperscript{521}

Thus, to the extent that items are semantically similar, they activate similar sets of units and to the extent that they are dissimilar, they activate different sets of units; “the vehicles of the several representations (activation patterns) will be similar to one another in ways which echo the semantic similarity of the cases”\textsuperscript{522}. Each unit may, for clarity, be thought of a kind of ‘sub-symbol’ corresponding to that sub-feature it picks out; however, even this notion – although helpful in understanding the nature of these networks – is misleading, as we shall discover below. There is thus a natural parallel between items being represented and the resources used to represent those items, and this is what is meant by ‘internally structured’ and ‘non-arbitrary’; “The semantic (broadly understood) similarity between representational \textit{contents} is echoed as a similarity between representational \textit{vehicles}. Within such a scheme, the representation of individual items is \textit{nonarbitrary}.”\textsuperscript{523} The status of the sub-symbols, rather then the symbols themselves, now takes center stage.

Smolensky’s coffee example\textsuperscript{524} is standard fare here, as is his vector analysis\textsuperscript{525}. ‘Coffee’ is never presented to the system as a bare particular, it is always ‘in a cup’ or as a ‘can with coffee’ or as a ‘brown liquid with burnt odour’, etc., and units (or groups of units) are associated with “code for [these] various features of coffee-involving scenarios”\textsuperscript{526}. The question is “What, then, constitutes the network’s representation of the conceptual constituent ‘coffee’?”\textsuperscript{527}. Taking some context like ‘cup with coffee’ and subtracting the ‘cup’ part (see Pylyshyn in MacDonald and MacDonald, 1995) does not work since it leaves us with “a number of contextually biased items, such as ‘brown liquid contacting porcelain’”\textsuperscript{528}, and similar biases result from similar moves. The representation of coffee is said to be context-sensitive in the sense that the form the representation takes is sensitive to what context the coffee is presented within. It seems there may be “no context-independent, core representation of coffee. Instead, there could be a variety of states linked merely by a relation of family resemblance.”\textsuperscript{529}

\textsuperscript{521} Clark, 1993, pg. 19
\textsuperscript{522} Clark, 1993, pg. 23
\textsuperscript{523} Clark, 1993, pg. 19
\textsuperscript{524} See Smolensky, 1991.
\textsuperscript{525} For a detailed treatment of Smolensky and vector analysis in particular, see Rand (2000).
\textsuperscript{526} Clark, 1993, pg. 24
\textsuperscript{527} Clark, 1993, pg. 24
\textsuperscript{528} Clark, 1993, pg. 24
\textsuperscript{529} Clark, 1993, pg. 24
What we appear to have are any number of sub-conceptual patterns of activity (line shapes in the case of the lexical system above, for example) corresponding to microfeatures of the semantic items in question (the letter shapes), the summation of which appears to correspond with the conceptual item (the letters themselves). Hence, large-scale patterns of activity correspond to semantic items, and sub-patterns to the microfeatures of which those items are composed. This assignment of activity to microfeatures, which together make up the semantic items, is the semantic metric at work. A relevant question now relates to the status of the microfeatures – *are they not symbols in their own right, albeit symbols at a lower level of analysis?*

The natural criticism of this analysis, then, is to ask whether one has not simply shifted the symbolic structure of the computing to a lower level - even if the microfeatures present at the lower level may not be understood, or be practically interpretable, by *post hoc* analysis. If that is the case, then surely the network is really just implementing a Turing machine at a different level of analysis – at the level of the letterforms, rather than whole letters, for example. A syntactic-semantic link would still hold - namely that between the microfeatures and the nodes - which is the cornerstone of the LOT hypothesis (see Chapter 4). It may an interesting feature of the network that the same resources are used for multiple representations, but so what? If the nodes are to be regarded as encoding non-arbitrary features of the inputs – albeit often some unknown but still definite microfeature - then what makes the claim that there is something interestingly different happening here, stand? Clark; “In what sense, if any, does the idea of a subsymbol or a subconceptual constituent amount to anything more interesting than the idea of smaller symbols?”

The answer lies in the sort of context-sensitivity at play, and in the sort of evolution that takes place in the network’s development.

### 5.3.2 Intrinsic (or Rampant) Context Sensitivity and Dynamic Structures

Context-sensitivity can be characterized in a number of ways. Clark introduces the notion as the identification of a semantic item within a varying *linguistic* context “*taken in reverse*”; a network

---

529 Clark, 1993, pg. 24
530 It should be noted, however, that these ‘toy’ examples are somewhat misleading in the sense that it is implied that the microfeatures are well understood (above they are the shapes of parts of the letters, below they are the ways in which coffee may be presented), but it should be noted that in real systems - in which there may be thousands (artificial systems) or millions (the brain) of units – there is no way of knowing what microfeatures the system has selected during training.
531 Clark, 1993, pg. 26
532 Clark, 1993, pg. 23
is presented with a label at the output layer\textsuperscript{533} and fixes on some relevant token of that semantic item that it had previously been exposed to. Thus “a net exposed to a label like ‘dog’ will need to fix on one of the several inner states which are associated with the label. To do so, the net relies on contextual information … [for example, see McClelland and Kawamoto (1986)] … in the context of ‘wore a woolly jacket’ [it] might be driven to a position in the overall representational space appropriate to a poodle feature complex\textsuperscript{534} (I take it that without additional context provided in the form of language-like labels, the network simply produces the prototype\textsuperscript{535}). Although a network’s ability to auto-associate relevant items based on sensitivity to the language surrounding a label is one of the most interesting and profound aspects of C-Nets, it is not the characteristic I shall focus on here. Instead, I shall continue to view the process of primary interest as one of categorization – since it most closely corresponds to the pattern-recognitional abilities associated with intuition – and context-sensitivity as being centered on the semantic metric and its relationship with some perceptual (or conceptual) field of data.

Context sensitivity for me, then, rests on the idea that the occurrent signal as a large-scale pattern of activity in the network is composed of any number of sub-patterns of activity, each associated with some (perhaps unknown) microfeature of the semantic item in question, and any interpretable large-scale pattern can thus be multiply instantiated depending on the context of presentation to the network. The representation of an ‘A’ in the lexical-identification network above is dependent on the form of the ‘A’ presented (in what font, for example) and is thus sensitive to the context in which the letter is embedded. The signal itself contains information pertinent to the context in which the semantic item is to be interpreted and so the context in such a C-Net is embedded within and is part and parcel of the representation, rather than in some number of additional symbols surrounding that representation; “in the symbolic paradigm the context of a symbol is manifest around it and consists of other symbols; in the subsymbolic paradigm the content of a symbol is manifest inside it, and consists of subsymbols.”\textsuperscript{536} The question remains as to status of these subsymbols, and whether or not they can be viewed as independently-defined syntactic items in their own right.

Before looking at the question of individual elements, it is important to raise the stakes a little. The context-sensitivity of the networks discussed so far has no real temporal component, other than as a

\textsuperscript{533} Keep in mind it is taken as an empirical fact that the C-Nets of interest are capable of operating in reverse in just this way – they are bi-directional.
\textsuperscript{534} Clark, 1993, pg. 23
\textsuperscript{535} See Section 5.3.1
\textsuperscript{536} Smolensky, 1988, pg. 17
variable over which the signal varies. However, many C-Nets of interest are known as ‘recurrant’ networks, in which the current processing (at a hidden layer, for example) is connected to a ‘context layer’ which stores some signal from a previous cycle of signal processing. Thus, the current hidden layer may be directly influenced by past system behaviour; context-sensitivity may then be extended and given an additional temporal dimension. A very simple example of such a network is Elman’s word categorization network (see Elman, 1991) – the task of which is to identify the class of word that should come next in a grammatically correct sequence - in which the context-layer stores the contents of previous signal’s hidden layer. Here …

… it is somewhat misleading to speak of the hidden unit representations as word representations in the conventional sense, since these patterns also reflect the prior context. As a result it is literally the case that every occurrence of a lexical item has a separate internal representation. … The representations are not propositional and their information content changes constantly over time in accord with the demands of the current task.\(^{537}\) [emphasis mine]

One may imaginatively expand on this simple architecture to envision deeply and multiply intertwined networks and layers, increasing this temporal context-dependence in scope to the extent that no signal may be interpreted without examining large portions of the rest of the network’s current activity and large swaths of temporal activity. It seems intuitive that such imaginative extrapolation is called for if we are to take seriously the fact that it is our brains that are of real interest here. Hence, no representing signal may be defined independently of the microfeatures corresponding to the nodes on which it is present, other signals within the network, as well as past activity within some portions of the network. Again – the question is whether such analysis bottoms out in context-free elements – albeit now in some unfathomably complex manner.

Following Clark, I argue that such systems do not, that the context-dependence of these systems is (what Clark calls) ‘rampant’, and there is no sense in which such systems bottom out at independent, context-free elements; C-Nets are “context sensitive ‘all the way down’”\(^{538}\). Although this claim may seem at first blush to be “nearly unintelligible”\(^{539}\) – it is only given the Turing paradigm of representations as static, syntactic physical structures that persist over time as physical entities that such a claim seems logically inconsistent. Given a new computing paradigm

---

\(^{537}\) Elman, 1991, pp. 353, 378
\(^{538}\) Clark, 1993, pp. 28-29
\(^{539}\) Clark, 1993, pg. 30
that is dynamic, rather than static, in nature, such a picture can be made sense of; indeed, it will be argued to be both consistent and true.

One must first ask: what is a context-free atom? What is this computational item that we are searching for and what is its relationship to those microfeatures that are meant to end the descent into smaller and smaller sub-symbols and sub-features? As we have seen in Chapter 4, this notion can be made quite precise within the LOT paradigm, and Clark’s definition certainly seems to be something Fodor would be happy with:

Something is a ‘context-free atom’ if (a) it is a syntactic item (ie, one that can be individuated purely by its nonsemantic properties) and (b) it plays a fixed representational role in a symbol system … Fodor’s innate Language of Thought posits a set of innate representational items in just this sense … Correlatively, consider the very idea of a microfeature. This can, it seems, be nothing but the idea of a context-free atom of the kind just described. What else could it be?\textsuperscript{540}

Patterns of activity – the propagating signals themselves – have been most closely associated with representational items (and indeed, I shall endorse the view that the propagating signal is the item most closely associated with traditional representational vehicles): perhaps these could be regarded as composed of context-free atoms. On this view, however, any particular pattern (or sub-pattern) is “affected by current context (that is, they will be slightly varied in ways dictated by other current activity in the system)”\textsuperscript{541} as well as – on my view – by the mode of presentation to the network. Such patterns of activity are indeed structured - since the semantic metric ensures that they are “nonarbitrarily constructed”\textsuperscript{542} - but the context-sensitivity of these signals is deeply embedded within that structure. Such an item does not fill the role of context-free atom.

Where else, then, might one find these context-free atoms? In the simplistic interpretation of C-Nets that I have so scrupulously tried to avoid, it is true that single units or patterns of units may be interpreted as such atoms. But in the C-Nets of interest I have taken great pains to emphasize why such an interpretation does not hold. The semantic metric imposes a set of micro-features on the inputs, which may indeed be associated with individual units or patterns of units for any single semantic item – but the real point of superpositional storage is that any number of semantic items may be coded for such a network on the same set of computing resources, and so any particular

\textsuperscript{540} Clark, 1993, pg. 31
\textsuperscript{541} Clark, 1993, pg. 32
subset of units plays multiple roles and cannot be identified with any single micro-feature\textsuperscript{543}. These context-free atoms are, on the traditional view, “syntactically defined item[s] taken as the persisting vehicle[s] of some single content”\textsuperscript{544} and such items “simply do not exist”\textsuperscript{545} in these sorts of C-Nets. To emphasize; “To justify the claim that there are context-free atoms acting as the representational baseline of such systems, we would need to discover syntactic structures which persist unaltered and carry a fixed content.”\textsuperscript{546} The only fixed structures here are the nodes and weights, but the problem we’ve encountered is that they “contribute[s] to multiple representations”\textsuperscript{547}.

The problem, according to Clark, is that trying to understand C-Nets in terms of symbolic computing “leads us to make the crucial error of trying to understand an essentially dynamic approach in essentially static terms”\textsuperscript{548}. Static structures of the sort sought by the traditionalist are simply not to be found; what these networks do is respond to a stimulus appropriately by generating signals conditioned by the ‘knowledge’ that is instantiated by the cumulative set of nodes, connections and weights. The ‘knowledge’ may be static, but there is nothing explicit about the semantic content of that static structure\textsuperscript{549}, for (as we shall see) it must be exposed to some appropriate activation signal for any meaningful structure to be made manifest. Clark; “Symbols, insofar as we can use such talk at all, exist in these systems only as a present response to an input of endogenous or exogenous origin. And these transient constructs internally reflect the local context in their very structure.”\textsuperscript{550} C-Nets are, then, essentially dynamic in their operation, and any search for static, interpretable items is doomed; “the idea of a symbol as both a bearer of some fixed content and a persistent inner item over which computational processes can be defined is not applicable to these models.”\textsuperscript{551} It is an artifact of the traditional view that context-free atoms are to be sought at all.

\textsuperscript{542} Clark, 1993, pg. 33
\textsuperscript{543} While it may be true that specialized areas of the brain, such as the visual regions v1 to v4, are specialized to the degree that well-defined functions are performed over well-defined inputs, even granting this claim does not guarantee the sort of node/micro-feature identity required by the symbolic paradigm.

\textsuperscript{544} Clark, 1993, pg. 32
\textsuperscript{545} Clark, 1993, pg. 32
\textsuperscript{546} Clark, 1993, pg. 32
\textsuperscript{547} Clark, 1993, pg. 32
\textsuperscript{548} Clark, 1993, pg. 32
\textsuperscript{549} See Section 5.4.3 and 5.4.4 for an analysis of ‘knowledge’ and ‘explicit’, on this view
\textsuperscript{550} Clark, 1993, pg. 32
\textsuperscript{551} Clark, 1993, pg. 33
5.3.3 Strong Representational Change; Code and Process

A key component of syntactically driven computing systems is the fact that the store-house of symbols in a given system is fixed; artificial systems may have more and more symbols added, of course, as a matter of design, but we saw in Fodor’s paradigmatic LOT thesis the process of learning new concepts – presented as a kind of generation and testing of hypotheses about the contents of the concept – ultimately required Fodor to commit to a kind of massive nativism\(^5\). This was because the conceptual resources required by, and constituting, the new concept would already have to be present in order to allow for the initial recognition to be a recognition that \(P\) (given \(P\) is the concept in question). Concept acquisition on Fodor’s view then, “does not result in any increase in the representational power of the system. To learn the ‘new’ concept at all, on the generate-and-test model, you must already possess the representational resources to express its content.”\(^5\) Representational change, then, viewed as the evolving capacity of the system to represent aspects of the world is fundamentally conservative in that “the products of such change … necessarily fall… within the expressive scope of the original representational base … it depicts rational representational change as limited by a preceding and representational base.”\(^5\) The environment is involved primarily as “a source of triggering experiences.”\(^5\) I follow Clark in calling this ‘weak representational change’.

This is in stark contrast to the C-Nets under consideration. Putting aside for a moment questions as to how and to what extent C-Nets are capable of instantiating some sort of native capacities\(^5\), the main point of contrast I want to bring out involves the role of the environment in shaping the system and altering its representational capacities. It is by having a training corpus presented to the network, along with some learning mechanism to alter the weights, that a network develops anything that may be considered a candidate for semantic or representational interpretation. The only candidates for representation are the patterns of activity in the network that comprise an output layer, and the development of new patterns of activity as the C-Net evolves and learns is fundamental to how these networks are to be understood.

\(^5\) See Section 4.2.4.2
\(^5\) Clark, 1993, pg. 34
\(^5\) Clark, 1993, pg. 35
\(^5\) Clark, 1993, pg. 35
\(^5\) I note now here that there are all sorts of interesting ways to view innateness from a C-Net perspective, as initial architectures and weight/connection settings are quite amenable to filling this role. In a C-Net, the only candidates for nativity are nodes and connections, and as we have seen, these cannot be viewed as instantiating semantic content in and of themselves. Note that systems that start from random weights and connections (such as NETTalk), that
C-Nets are strongly ‘data driven’, in the sense that it is the presentation of various inputs that shapes the very processing of the system itself. Indeed, not only does the representational base change as the system’s behaviour changes, it is an empirical fact about C-Nets that changes in behaviour can happen quite suddenly and dramatically; “continued training can bring about qualitative changes in the performance of a network … Critical mass effects are … to be expected, for we are dealing with highly data-driven systems whose processing profiles are an immediate consequence of their state of superposed stored knowledge”\(^5\). The changes in behaviour of these non-linear systems can change quite abruptly and that change in behaviour corresponds quite directly to a change in representational capacities. The point here is that the representational capacities of a C-Net are not fixed, but change as the processing profile of the network itself changes.

Given that the set of weights may be considered as comprising the systems ‘knowledge’, there is a “deep interpenetration of knowledge and processing characteristics which is fundamental to the connectionist approaches”\(^5\). Since new knowledge acquired by C-Nets – in the form of newly-developed capacities to respond differentially to additional items of interest – comes in the form of altering a given set of weights, and given that the set of weights is the source of the ‘processing’ of the system, “Changes in the knowledge base and in the processing characteristics go hand in hand.” The input corpus, the processing profile and the representational capacities of the system are all deeply intertwined, and since the input corpus is by no means fixed – it comes from encounters with the world - neither is the processing profile or representational capacity. McClelland, Rumelhart and Hinton;

“The representation of the knowledge is set up in such a way that the knowledge necessarily influences the course of processing. Using knowledge in processing is no longer a matter of finding the relevant information in memory and bringing it to bear: it is part and parcel of the processing itself.”\(^5\)

\(^5\) Clark, 1993, pp. 38-39
\(^5\) Clark, 1993, pg. 39
\(^5\) McClelland, Rumelhart & Hinton, 1986, pg. 32
Thus, far from the Fodorian picture of a fixed representational base and massive nativism, C-Nets have little (if any) innate symbolic content and evolve over time and fundamentally change their processing profile and representational capacities dependent on exogenous sources (the world).

5.3.4 Concluding Remarks

C-Nets are processing systems that are quite different from traditional Turing machines. The most fundamental difference is that there are no syntactic structures to be found in a C-Net that would correspond to, or play the role of, symbols over which processing is defined. They also do not have a fixed representational base, rather, there is a deep interpenetration of data and processing, as the processing changes as new data is presented and absorbed into the set of superposed weights, and the representational base correspondingly changes.

C-Nets are a dynamic, data-driven, highly complex and non-linear stimulus-response mechanism, the output of which is the best contender for context-free representational content - and I do mean contender since context-free content is not clearly existent at all, without the sorts of qualifications I provide in my account of pragmatic conceptualism below. That output is not a physically persisting, static structure, but something that is transiently produced given some input stimulus, and is highly context-dependent though structured none-the-less.

C-Nets, characterized in this way, undoubtedly have much ballyhooed shortcomings - in particular an inability to handle conceptual content in a way that satisfies some sort of compositionality constraint, which was presented as an explanatory strength of the LOT thesis – but they are certainly not to be regarded as Turing machines and as such, are not susceptible to the criticisms of Chapters 2 and 3. They are to be regarded as highly sensitive pattern-recognition devices. How C-Nets may overcome their short-comings is addressed below, but ultimately, I put aside those questions by postulating that it is the role of C-Nets as implementers (or approximations) of Turing machines that solves that riddle. The special computing resources that C-Nets bring to bear can be supplementary to whatever other cognitive abilities we may have, although there certainly remain interesting questions regarding how C-Nets instantiate those (Turing) properties as well.
5.4 That C-Nets are Not Turing Machine – What Follows?

Here, I develop some implications of the view of C-Nets developed thus far, working my way to a final characterization of both concepts and intuition from a C-Net perspective. I start with the explanatory relationship that holds between the intentionally-described state of affairs and the underlying cognitive structure, which requires - in the C-Net environment - a kind of post-hoc analysis. C-Nets are not understood in the same way a classic LOT system would be, and the sorts of mathematical analysis associated with attempts to understand the operations of C-Nets often lead to accusations that such analyses are non-explanatory, in that they are post-hoc in nature. In arguing against these criticisms I urge that the proper way to view such networks is from an externalist perspective – which provides for a way to flesh out the semantic content of concepts - although I simultaneously deny the need to be a behaviourist in advocating a primary functional role to consciousness.

Extending this position, I outline my view of concepts and set out the differentiation between concepts\textsubscript{mental} and concepts\textsubscript{public}. Concepts\textsubscript{mental} are, on this view, counter-factually defined sets of transient signals that are highly individualistic and context-dependent, while concepts\textsubscript{public} are more like the expressions of competence demonstrated by a network. I further contrast this view against the LOT thesis by articulating a distinction between implicit and explicit knowledge structures that incorporates the processing of data itself, as part and parcel of that distinction. C-Nets can be viewed as containing explicit structure (even though highly context-dependent) although ‘explicit’ needs to be unhinged from an intuitively textual feel with which it is normally associated. From that perspective, weights and connections can be regarded as a kind of accumulated knowledge, which – finally - can find expression as a spontaneous judgment of fact or appropriate recognition/classification of a pattern.

Intuition, then, is given as C-Nets response to some activating signal – conceptual or perceptual – that is shaped by previous encounters with relevantly similar phenomenon. Intuitions may be regarded as elements of the set of signals that comprise a concept.

5.4.1 Explanatory Inversion and Post-Hoc Analysis
One of the strengths of the LOT thesis is that it is capable of explaining psychological phenomena. The reason you go to the refrigerator to get a carrot is because you want a carrot and believe a carrot is in the refrigerator. Each of these beliefs and desires is suitably encoded in your brain, via a syntactic structure that varies with the content of those beliefs and desires, and that syntactic structure plays the appropriate causal role in delivering you to the refrigerator and opening the door. Thus, the level of abstract theorizing – the psychological description of your intentional states in everyday language – bears a direct relation to the causal, computational processes that instantiate those intentional entities.

Recall Marr’s three distinct levels of explanation, which are normally applied to cognition *writ large* (and not just vision, which was Marr’s particular target). There is a top-level (Level 1) comprised of the abstract formulation of what task is being performed – often called a competence theory - (here, the folk psychological description), a lower-level (Level 2) that is comprised of “a particular algorithm for carrying out the computation”560 and finally the implementation level (Level 3) composed of syntactic bits and bobs or nodes and connections. In the classic LOT paradigm, there is, in Dennett’s words “a triumphant cascade through Marr’s three levels”561. Each level is distinct, but there is an isomorphic relation that holds from Level 1 right through Level 3. This point should be clear from Chapter 4. Indeed, an IDR adherent who adopted the classical view need only to claim that what is being sought by the use of intuition is access to the Level 3 component(s) corresponding to the Level 1 term (certainly an IDR adherent is free to remain neutral with respect to issues of implementation, but committing to the classical cascade in this way is certainly one way to deliver the semantic access she seeks).

A connectionist has a different view of things. The relationship that holds between the three levels differs mainly in two ways. First, the direction of influence is two-way; that is, not only does Level 1 constrain the equivalence class of possible Level 3 implementations, but the way in which Level 3 gets the job done has implications for how Level 1 is best characterized. This is really what I mean when I say “implementation matters”. I address this matter in the sections that follow. Of current concern is the fact that, under the purview of a connectionist, the way in which Level 1 is to be characterized as a competence theory is one which only roughly describes the cognitive activities implemented in Level 3. Level 3 captures the cognition that actually occurs (via the non-linear dynamic equation), and Level 1 is regarded as a mere *approximation* of that activity; Level 3 is far

---

560 Clark, 1993, pg. 44
561 Dennett, 1998, pg. 227
more nuanced and exhibits far more fine-grained behaviour than is captured by the Level 1 explanation. It is this second difference upon which I focus here.

In denying that there exists a ‘classical cascade’, the connectionist “gains an explanatory obligation”\(^5\); hence, the ‘explanatory inversion’. In the classical case, Level 1 acts to explain what is happening ‘down below’ (which fixes the equivalence class of appropriate implementations), whereas the problem for the connectionist is that “the unavailability of the classical competence theory thus threatens to render connectionist models non-explanatory in a very deep sense”\(^5\). Level 1 has become the *explananda*, rather than the *explanandum*. A connectionist needs to explain how the structure associated with the intentionally described competence theory emerges from the soupy mix of nodes, connections and signals which comprise Level 3.

In order to fill this explanatory burden, a number of mathematical tools have been developed to provide a *post-hoc* analysis of the dynamics of a network. Cluster analysis, for example, enables one to visualize the topological space that separates the various objects that a network classifies. Vector analysis (see Smolensky, 1998, Rand 2000) performs much the same function, but allows for arbitrarily many semantic interpretations of the same data. More advanced techniques (see Ripley, 1996, McClelland & Rumelhart, 1986) allow for that topological space to be described in dynamic terms, a mathematical model that examines changes in state-spaces over time so that areas of *stability* can be identified. In each of these techniques, the non-linear dynamic equation is taken as the locus of activity to be described, and in all of these techniques anything remotely associated with symbolic structure are “unearthed [as] abstractions over sets of much more highly structured distributed representations … [and] … will offer at best a rough guide to some facets of the system’s behaviour”\(^5\). In other words, when we seek to understand the operation of these networks, we can only do so by modeling their behaviour in ways that brings out approximations to the structure we are seeking, and different techniques will necessarily perform those approximations in different ways.

So it turns out to be a matter of practice that the structures involved in such a competence theory like the intentional domain, only emerge to view – and only emerge *approximately* - when we seek to describe the operation of the network in a way we can understand. We can only understand that operation by the system’s behaviour, and the mathematical models describe that behaviour, and we understand the relevant behaviour using intentional language. This is why “[t]he connectionist

\(^{562}\) Clark, 1993, pg. 67  
\(^{563}\) Clark, 1993, pg. 49  
\(^{564}\) Clark, 1993, pg. 61
takes … an externalist attitude toward [competence-theoretic stories], seeing them as descriptive of what gets done (that is, of the abilities of the system).”

Using these post-hoc techniques, “a competence theory emerges as a body of laws which serve to pick out the states into which the system will settle under certain ideal conditions … [it] is a kind of grammar which fixes on certain stable states of the system … [and they] capture (at best) some central patterns in the processing and … do not exhaust semantic nuances.” So Level 1 is descriptively adequate in helping us understand what a network is up to, but it does not mirror the cognitive dynamics in the same way the classical theory does.

In adopting this externalist view toward competence theories, the connectionist is dangerously close to behaviourism on the one hand and a completely private language on the other. Behaviourism would follow if the only mental content that can be attributed to the networks is visible at the level of behaviour, and a completely private language would follow if the dynamics of each network were distinct. Of what structures might a connectionist avail herself to escape such opposing but equally barren and hostile territories?

5.4.2 Prototypes & Statistics, Theory & Knowledge, Intuition

In attempting to make visible the structures necessary to both understand the C-Nets actual operation and to explain the emergence of conceptual content (which I take it will generally stand in for all the requisite intentional and psychological structure we are seeking), we may be tempted to posit that these emergent, approximate abstractions really are just the symbols we seek after all. This will not quite do, however, as this temptation “can lead us to hallucinate familiar, manipulable symbolic items where none exist.” There is structure, but it is not the symbolic structure the classicist demands.

The strength of C-Nets most universally understood in being brought out by mathematical treatments is the ability to generate a statistical central tendency of the exemplars – I take this to be well-established. There are two interesting points that follow from the establishment of the generation of a statistical central tendency: the first is the relationship of the statistical central tendency to more traditional prototype theories, and the second is the implied unavailability on the statistical account of theoretical terms in the generation of concepts in category judgments.

565 Clark, 1993, pg. 47
566 Clark, 1993, pp. 48,49
567 Clark, 1993, pg. 42
With regard to the first, I shall argue that the statistical analysis offers an even more psychologically suggestive explanation (than prototype theory) for certain judgments; recall that prototype theories were first suggested in order to account for typicality and scalar effects\(^{568}\) in category judgments, in terms of both speed of response to typical members of a class and graded class membership, and in that sense prototype theorists gained traction by the psychologically suggestive nature of their explanation. I shall argue that C-Nets inherit the psychologically suggestive advantages of prototype theories, but do not likewise inherit what is commonly regarded as a drawback of such theories, that of prototype combination.

With regard to the second, straightforward statistical central tendencies do, at first glance, seem to pose a serious problem to more theory-laden (and especially novel) classificatory judgments; “A variety of difficulties face such purely statistical approaches. The major drawback is that such approaches seem to ignore the role of theoretical knowledge in informing judgments of category membership.”\(^{569}\) Clearly a method of generating theoretical terms in the C-Net economy is required; “… we are capable of forming abstract concepts, and concepts of unencountered (and unencounterable) situations and objects. Such cases cannot be explained by the extraction of a central tendency from a class of sound exemplars.”\(^{570}\) I shall argue that theoretical terms are not a problem for the connectionist, hence theoreticity is a “pseudo-problem”\(^{571}\).

5.4.2.1 Typicality Judgments; Prototype Theory & The Statistical Central Tendency

It is well understood that C-Nets quite naturally extract statistical relationships among the training corpus, and here I argue that – even though that statistical relationship can in some ways be well-described as a kind of prototype extraction – it is a mistake to think of a prototype as being literally deployed. Aside from the general dynamical consideration that there are no context-independent static features to be found in such a network of which the prototype is supposed to be constructed, I

\(^{568}\) For a full treatment see Rosch (1975, 1978) and Lakoff & Johnson (1999). Typicality effects are such that some elements are considered more typical of some class membership than another (a robin is more typical bird than an ostrich, and a bird is more like an animal than a flea), and the scalar effect extends that notion in that “we often judge an instance P as falling under a concept Q to a greater or lesser extent” (Clark, pg. 88), which translates on the prototype theory to a distance from the prototype. Further psychological evidence of the scalar effect is generated by measuring the speed of response to a more or less typical example; a faster response is generally regarded to be evidence of less ‘cognitive distance’ from the constructed prototype.

\(^{569}\) Clark, 1993, pg. 90

\(^{570}\) Clark, 1993, pg. 90

\(^{571}\) Clark, 1993, pg. 103
provide two other considerations. First, I argue that what was meant to be an explanatory strength of prototype theories – that it better accounts for typicality judgments than other conceptual schemes – is actually better seen as a strength of the C-Net statistical account since the C-Net account is in a (slightly) better position to account for conceptual combination. Second, I provide some empirical evidence that positing prototype construction – or more generally the construction of mental representations of deep structure common among an input corpus - is not the best explanation for various category judgments.

I take it that typicality judgments are well-understood, at least insofar as how they are described from an empirical perspective. Typicality judgments are the phenomena that some members of a class are judged to be more typical of that class than others, and further, that there is a scaling effect to such judgments - that is, members are graded so as to be more or less typical of a category. For example, a cat is judged to be a more typical member of the category ‘animal’ than a flea. Clark;

... typicality judgments are … surprisingly unstable across contexts … [and] such judgments are ubiquitous insofar as there seems to be no categories or concepts that are exempt from the scalar grading effect. Everything from basic taxonomic concepts (‘fruit’, ‘furniture’) through abstract and formal concepts (‘odd number’, ‘square’) to ad hoc categories (‘ways to escape being killed by the Mafia’) and on to goal-directed categories (‘things to eat on a diet’) has been shown to display graded structure … and these gradations are psychologically potent insofar as they are also correlated with response times (the more typical cases are more quickly categorized) and with ease of acquisition. Typicality judgments, then, must be taken seriously.

Prototype theory was motivated, to a large extent, by a desire to explain this phenomenon, particularly the ‘psychologically potent’ part that correspond to response times. Judgments of category membership are thought to be made by comparing the object directly to a constructed mental prototype, or typical class member. The scalar effect relates to the distance from the prototype, and ‘distance’ here is meant to capture the relevant semantic difference. How that difference is measured depends on your favourite flavour of prototype theory - and so it can be defined in a variety of ways - but typically it is measured as some additive value over a set of weighted individual features that make up the prototype (the weight corresponding to the

573 Clark, 1993, pp. 91-92
importance of that feature to the category). The details of any individual prototype theory is not important here - the central point is that there is a static, mental construct consisting of features or properties, which acts as a point of comparison to the potential class members being judged. Time of response indicates cognitive work, which in turn is meant to indicate the semantic distance from the potential member to the prototype. The empirical fact that there are faster response times for more typical members of a given category is meant to count as evidence of the psychological reality of the prototype itself.

This explanatory strength comes at a cost: prototype theories are saddled with the charge that there is “no good theory of prototype combination”\(^{574}\) (a charge we shall see the C-Net variant is able to dodge, albeit in quite restricted domains; general conceptual recombination and redeployment will remain a problem). A theory of prototype combination is required - on the view that prototypes are literal constructs - in order to account for the sorts of ad hoc categories referred to above (as in the case of judging some object to be an example of the category ‘things I’d take on a winter trip to the bottom of the sea’), and certainly for highly theoretical judgments involving novel thought experiments (like the case of asking someone whether ‘is this incredibly imaginative and complex case of being hit on the head and accidentally knowing the temperature outside a genuine case of knowledge?’). Of course, an alternative to prototype combination is to assert that the various prototypes already exist. It is not difficult to see, however, that the problem of massive nativism – quite similar to that which plagued Fodor’s account - follows; and it seems reasonable to respond “[T]here just can’t be a distinct prototype for every complex expression.”. On the prototype view, then, some new prototype would need to be constructed on-the-hoof to deal with these sorts of ad hoc and highly theoretical situations.

I shall assume that the problem of prototype combination is well-established (see Fodor 1981, Prinz 2002); combining the prototypes of the elements of some complex expression results in “an incoherent mess”\(^{575}\). Standard fare include such examples as the expression ‘pet fish’ - try to imagine what the combinatorial prototype is like if it is meant to be a combination of a prototypical ‘fish’ and a prototypical ‘pet’. Does it really look like a goldfish? A prototypical Harvard carpenter is supposed to be non-materialistic – do either prototype elements contain this property? Here, there seems to be a kind of emergent property associated with the phrase (pet fish, Harvard carpenter) that cannot be regarded as composed of elements corresponding with the prototypes associated with the lexical components.

\(^{574}\) Clark, 1993. pg. 106
\(^{575}\) Clark, 1993. pg. 106
C-Nets fare much better, albeit in a restricted sense. Recall that the semantic metric acts as a sort of state-space into which objects are placed relative to each other according to their degree of similarity. The operation of the semantic metric implies that “features common to the exemplars become strongly associated” since they activate common connections, and this naturally leads to a system which will “extract the statistical central tendency of the exemplars …[and] it will uncover which sets of features are most commonly present in the learning set. It will also learn commonly occurring groupings of features.” Since superpositional storage requires that existing resources must be modified to encode new information, then those features which occur most frequently among the input corpus become “highly marked and mutually associated.” That is, the connections associated with those commonly co-occurring features are both strong and “become encoded in such a way that activation of the resources encoding one such feature will promote activation of the other.” This leads quite naturally to a kind of prototype extraction, since both statistical central tendencies as well as ‘stereotypical feature sets’ are quite naturally encoded in the set of weights and connections. This is not to say that identifiable and discrete feature sets are associated with particular nodes and connections, rather it is to say that the network acts as if it were generating those feature sets as a kind of statistical response.

A key difference from prototype theory, however, is that the prototype need never be actually encountered – it is a generated ‘fiction’ of the network. It is useful to speak of a prototype, but “the prototype will be nothing but the center of a statistically driven state space in which the overall tendencies of attribute co-occurrence (the ‘central tendency’ of the category members) fix the abstract prototype which organizes future judgments. [emphasis mine]” The ‘nothing but’ here is a bit of a double-edged sword. On the one hand, it is what enables the connectionist to dodge the ‘prototype combination’ bullet (as I shall show presently), but it is also what allows critics of C-Nets theory to level the accusation that C-Nets are incapable of dealing with theoretical terms in categorical judgments (which I shall address in the next section). With regard to prototype combination, the ‘nothing but’ [a statistical tendency] allows for a kind of mutual engagement between prototypes to take place, without having to commit to a full-blown, actual prototype combination. Clark; “the connectionist version of prototype representation … is not a simple property list but is rather a complex of interlocking microrepresentational elements well suited to

\[576\] Clark, 1993, pg. 20
\[577\] Clark, 1993, pg. 20
\[578\] Clark, 1993, pg. 20
\[579\] Clark, 1993, pg. 21
\[580\] Clark, 1993, pg. 90
engage with certain other complexes. [emphasis mine]581 Here, Clark is arguing that C-Nets are well-suited to combine different sub-complexes in a way that is unsuited to actual prototypes.

The C-Net version of prototype combination is fairly straightforward “when the combined content lies within a single, pre-existing representational space”582, since the signals themselves are additive and the network forms a single semantic metric over which those signals may be regarded as having semantic content. C-Nets prototype combination does not consist in the “linear addition of the properties of each constituent prototype”583, but rather it is a more dynamic notion which I associate with a ‘cognitive forces’ view of mind; Horgan & Tienson characterize cognition in just these sorts of terms; “… the basis of cognitive processing is content-appropriate cognitive forces. … In general, content-relevant interaction of cognitive states is automatic for any two cognitive states than can be instantiated simultaneously in the same (sub)network”584

A conceptual combination involves “finding a midpoint in an existing representational space”585 and this is accomplished – depending on your favourite flavour of mathematical analysis – by some “delicate process of mutual activation and inhibition”586. One can clearly see the ‘cognitive forces’ view of mind coming into play here, with the notion of signals inhibiting and re-enforcing each other within some semantic metric. There remains, for Clark, a deep problem because this analysis holds only on a single network and so “the connectionist … lacks any general account of conceptual combination [emphasis mine]”587 since a general account for a theory of concepts requires that the conceptual content account for “the ability to entertain any one of an infinity of potential thoughts involving the systematic recombination of grasped meanings”588. I hold off on my solution to this problem until Section 5.4.4.

It seems clear, however, that typicality judgments can be better explained by a C-Net theory, than by a prototype theory since the C-Net theorist has some resources available to attack the recombination problem. The C-Net theory has all the psychological suggestiveness of the prototype theory, with one of the major theoretical concerns – if not out of the way – at least pushed to one side. Note that these sorts of judgments fit the profile of an intuitive judgment, since no features or properties of

581 Clark, 1993, pg. 107
582 Clark, 1993, pg. 112
583 Clark, 1993, pg. 107
585 Clark, 1993, pg. 107
586 Clark, 1993, pg. 107
587 Clark, 1993, pg. 112
588 Clark, 1993, pg. 108
the prototype need be instantiated and certainly need not be brought onto the stage of consciousness. Recall object recognition was classified as O-Intuition, an example of intuiting the obvious. To elevate the purview of these judgments to more interesting cases, however, we need to address the issue of the role of theory in categorization judgments. This is the topic of the next section.

First, though, I note there is empirical evidence that prototypes are not constructed and deployed in categorization tasks. Here I follow Whittlesea & Dorken (W&D, 1993). The question they pose is more general, and relates to whether positing the construction of any temporally enduring mental structure - matching that associated with the commonalities among the elements of a training set - is the best way to explain category judgments related to that set. They ask: do we generate an “abstraction of general deep structure” that encompasses the set of stimuli (which would include the structure of an prototype) or do we “simply code each exemplar as it is presented”? What W&D want to show is that “the deep-structural relations imposed by the grammar [the rule that generates the exemplars] are not represented in a precomputed format, but lie implicit in and distributed over the subjects’ knowledge of the exemplars generated by those rules …” In this case, the ‘grammar’ is presented in the form of a series of nonsense words in which the letters of successive cases are modified by a rule that relates each successive word, not to a central prototype, but to a semantic distance from the previous word. Subjects are to judge whether some novel example conforms to the underlying rule or not. Subjects were found to classify cases such that successive examples were judged in a way that gave greater weight to recently encountered examples than to that which would represent the extracted prototype of all examples encountered.

W&D conclude that “…there is no positive evidence of automatic, unconscious abstraction or coding of any type of structure. Instead, [they] conclude that memory performs and preserves whatever operations are functional in satisfying current demands, and accidentally becomes sensitive to underlying structure.” This is no proof, of course (perhaps prototypes are constantly being constructed on the fly, with greater emphasis placed on recent encounters; this is not a parsimonious explanation - and one rightly rejected by W&D - but logically possible none-the-less) but the emphasis on current operational demands does reflect the dynamic nature emphasized by the connectionists and is directly reflective of an on-going statistical central tendency that is weighted by the temporal occurrence of the items.

589 Whittlesea & Dorken, 1993, pg. 227
590 Whittlesea & Dorken, 1993, pg. 228
591 Whittlesea & Dorken, 1993, pg. 228
592 Whittlesea & Dorken, 1993, pg. 231
5.4.2.2 The Role of Theory & Knowledge

Clark: “A major drawback [to purely statistical approaches is that they] seem to ignore the role of theoretical knowledge in informing judgments of category membership … and it seems fairly obvious that at least some of our concepts must, in this sense, be driven by theory … such cases cannot be explained by extraction of a central tendency from a class of sound exemplars.” Clark, 1993, pg. 90 A typical example is that a skunk in a dog-costume remains a skunk, regardless of the perceptual change, hence the perceptual cues by themselves are insufficient to account for our judgments; a theoretical term like ‘has skunk parents’ surely plays a role in our forming judgments related to this concept. So much the case for theory w.r.t. skunks, so much greater the case for intentional terms like knowledge. I take this claim about the central role of theory in our judgments to be true, and it is to this problem that I now turn. The central question is two-fold: what is the nature of a theory of some domain, and how may such theoretical knowledge be made manifest in the weights-and-connections environment of a C-Net?

The proposal from Clark is to “push the simple idea that the theory of the domain might just be a set of weights in a connectionist network.” Clark, 1993, pp. 101-102 As I see it, this key move relates back to what I advocated at the start of this chapter – that in order to really give C-Nets their due, we must be able to imagine, not just single networks transforming a stream of incoming signals, but a network of networks in which a part of the input to any given network can consist of the output of other networks. In other words, signals may be pre-processed in any number of ways prior to being thought of as the primary resources undergoing some further transformation. Clark’s interpretation of this picture emerges as the claim that “we should insist that the connectionist is not committed to the use of only low-level perceptual features to define the state space.” Clark, 1993, pp. 101-102 Certainly, and I take it that the reason this is so is due to this possible layering of networks. Clark seems to rely on the same picture, although he (frustratingly) doesn’t explicitly endorse the ‘network-of-networks’ picture; “… presumably any realistic psychological model of the formation of a raccoon recognition space will allow as input any information to which a child is normally privy, including functional and genealogical information. … Once we allow a cascade of layers of units, and a wide variety of types of information as inputs, it seems quite possible to imagine multi-dimensional features spaces which do indeed rate ‘has raccoon parents’ as an important … condition of
raccoonhood.” Clark’s use of the term ‘information’ in this passage can only be given the weight he needs it to carry if there is some prior semantic appraisal of the signal. In a C-Net environment, that can only be as a signal transformed into an encoded output.

Is this really deploying a theory? In one sense, the answer must be no, for what we mean by that expression is often meant to imply that some sort of conscious and deep understanding is brought to bear on a given situation, can be consciously modified, etc. – and although there may be connectionist answers to how this is done, this is beyond my purview. However, if the deployment of a theory can be regarded as “evidence of a shift in judgments based on larger and more varied bodies of knowledge,” then the answer is clearly yes. Clark is also concerned with this narrower view; “… it is clear that any system which can use stored knowledge to reject or modify a judgment it would otherwise have made (on the basis of a perceptual input) counts as deploying some kind of theory of the domain.”

There are a number of ways in which theoretical information may be stored in a network – even on a single network view, and not the network-of-networks view I am advocating - and the capacity of non-linear networks to be made sensitive to arbitrarily subtle differences in inputs is one way to make it happen. Another way of putting this claim is to say that inputs that are very similar can be treated in very different ways, and the detectable difference can be articulated in a number of ways. It may even be temporal in nature, given a strong recurrent character of the network; prototypical sequences of events may be regarded as making contact with a causal story, for example. Given such ‘subtly present’ information as a part of the input, it becomes hard not to use the term ‘beliefs’ as being what modifies classification. Indeed, a lot of engineering work has been done on what are called ‘belief networks’. Ripley; “Other [supervised] methods incorporate non-numerical ‘real-world’ knowledge about the subject domain into the structure of the probability distributions. Such knowledge is often about causal relationships …” It should be noted that these beliefs are built into the system itself, and avoids the thorny problem of how an unsupervised network would form the proper ‘beliefs’. Clark admits that supervision avoids some tricky ground – “What is perhaps a little harder to see is how a network might discover such a normative feature [like ‘has raccoon parents’] for itself …” – but I have assumed that this is a practical and not a philosophical problem.

---

595 Clark, 1993, pg. 102
596 Clark, 1993, pg. 103
597 See Section 5.4.3 for an attempt.
598 Clark, 1993, pg. 104
599 Clark, 1993, pg. 104
Thus, Clark concludes that – even in the single network case – “the picture of pattern associators as indiscriminate sponges condemned to be sensitive to irrelevant and misleading correlations among the training data is too simplistic” and hence, the “problem of theoreticity is a pseudo-problem.” That being the case for single networks, it is surely all the more so for a psychologically realistic network-of-networks, in which theoretical terms as input can be quite easily accommodated; on the network-of-networks view, ‘having parents of the same species’ may be an associated signal from a parallel network that is triggered by all things animal. If single networks can help themselves to theoretical resources by virtue of being able to trigger on arbitrarily sensitive patterns in the input corpus, then inter-network signal passing just raises the theoretical stakes.

5.4.2.3 Explicit Knowledge Structures

On the view I am advocating, it is the set of weights and connections that is to be regarded as a kind of accumulated knowledge, with signals generated by those weights and connections playing the role of theoretical terms. If the set of weights and connections is to be thought of as a network’s accumulated knowledge, how then is that knowledge instantiated? Given that code and process – instantiated signals and the dynamic computing performed on those signals – are inextricably linked, how and where might be stored the individual facts which would normally be seen as comprising that knowledge? The criticism can be put thus: “insofar as [C-Nets] violate the code/process distinction, they cannot be seen as tokening anything explicitly.” A resulting problem, as I see it, is that knowledge - if only implicitly present in the network – cannot be given a very active role in my cogitations. Something stronger needs to be said, or ghosts of behaviourism past appear at the gates: without a notion of explicitness, it cannot be maintained that particular, isolable items of knowledge are causally implicated in belief generation, and the like. This problem is also related to problems of conceptual modularity and the accompanying accusations of causal holism and here I develop some of the resources required to counter that attack. I follow Clark in disambiguating the sense in which a mental data structure may be thought to be an explicit item on the C-Net view.

600 Ripley, 1996, pg. 243  
601 Clark, 1993, pg. 103  
602 Clark, 1993, pg. 103  
603 Clark, 1993, pg. 122  
604 See Section 5.4.3
Clark claims – and I agree - that traditional notions of explicitness are tightly bound up in our predilection to view information from a textual perspective, as “words in a text”, and this view permeates the classical view of computing. Following Kirsh, Clark identifies the assumptions underlying this view; information is traditionally thought to be explicit if it is structurally local, movable (it retains its meaning independent of spatial changes) and easily available. The first two properties we have encountered before as part and parcel of the classicist view, and their relevance to the notion of explicitness developed here shall be rejected. The last – ease of availability - is more interesting and gets much more to the heart of the matter.

A first counter-example will generate some (initial) intuitions (to be viewed with the requisite scepticism!) regarding the properties to be rejected; structural locality and movability. Encrypted information, it seems clear, should not count as explicit. A lot of processing is required to translate it into meaningful information and so it is certainly not easily available (if it were, there would not be any internet banking). Yet, it certainly would fit both of the properties we would like to reject: it is movable (again, home banking is built on this premise) and it is structurally local in the same way that is any classicist piece of data.

Now consider a second example. A piece of traditionally textual information is located somewhere in a massive set of encyclopedias (or better yet, stored somewhere on the internet), but for that set of books there is no index (and for the internet, no search engine). Is this item of information to be considered explicit? Here, the role of ease of availability plays a key role. If we reject that property as necessary for explicitness, then it seems we may say that the hidden information is explicit, but we have now lost the ability to deal with the encrypted example. If we retain it, then we can still encompass our intuitions about the encrypted example, but now we must say the hard-to-find information is not explicit. It seems to me that - traditional prejudices about the textual nature of explicit data structures aside – the hidden information can be quite readily thought of as not being explicit.

The upshot of all this is to say “that the theoretical core of the idea of explicitness should be the notion of easy usability of information”. The reason we are so taken with the structural notion explicitness, typified by the textual, is simply because such a form fills the ease of availability requirement for us “and we therefore tend to hallucinate that something like the structural forms of

---

605 Clark, 1993, pg. 122
606 Kirsh, 1991, pg. 350-358
natural language text is in fact essential to genuine explicitness. Our need for ease of availability requires text to be movable and structurally local, but – and this is a key move – these two properties are so defined only relative to us as processors of that information. Other processors may not require information to be presented in such a format, which leads to the following sort of proposal: Explicitness is only defined relative to some processor, the inclusion of which need render the information easily available (to that processor). So – relative to the human visual system – text is explicit.

Clark; “the extent to which context dependence defeats explicitness is relative to the ease with which context is taken into account by the processor”. His concern seems to be to account for the C-Nets ability to process “contextually nuanced data” in the form of the input signal – and I certainly take his point. Information that is context dependent (and hence implicit from some perspective) in raw perceptual data may be rendered explicit when the processor is a C-Net. However, the processor is rendering explicit only items in the incoming signal – the surrounding environment. I want to extend that idea from environmental signals to internal signals – signals within the network itself. I posit that a similar notion of explicitness can be brought to bear on the picture of the set of weights and connections as knowledge, in the sense that the network is capable of ‘reading’ off that knowledge itself, given some appropriate input signal. The line I’m pushing here is that just as text is informationally explicit to the human visual system, so too is the set of connections associated with accumulated knowledge informationally explicit to the network itself.

Here, we have the first glimpse of the functional role I shall assign to consciousness: consciousness allows, within the cognitive economy of the brain, for the implicit to be rendered explicit. Buried in this analysis is the notion of the network processing itself – and implicit to that is some sense of self-awareness – and I certainly admit that in order to make this notion of knowledge explicit I do rely on those sorts of resources. I now expand on just that notion, and define just what role consciousness must play, in my analysis of my theory of concepts: pragmatic conceptualism.

5.4.3 Pragmatic Conceptualism: A Theory of Concepts

I turn now to the question of what a concept might be, with C-Net goggles on. It is, after all, the concept of ‘knowledge’ (or some other term) that the intuitive judgments of concern are meant to

---

607 Clark, 1993, pg. 123
608 Clark, 1993, pg. 123
609 Clark, 1993, pg. 124
ferret out of our grey networks, according to the SPM. Let me first note, however, that this theory of concepts does not need to go through in its entirety in order for my characterization of intuitive judgments to go through, since that characterization is meant to stand on its own, and much of it is independent of the details of pragmatic conceptualism. This view of concepts is put forth in some detail in the interest of completeness, and to commit to a larger picture into which my notion of intuition falls.

First, I shall give my account of concepts, which – although deriving a lot of support from Clark – does depart from his account to some extent. Clark sees concepts as abilities, or more precisely, a bag of abilities – but I see them as having two modes which I am careful to differentiate; a counterfactually defined set of occurrent signals that play a causal role in mental life, and a publicly accessible, semantic mode. These two modes are linked by the role of consciousness. I am particularly concerned with articulating the role concepts might play in our mental lives, given an extended set of C-Net resources (to which I am willing to help myself). Second, I shall admit of some explanatory burdens unmet - as much as C-Net theory can contribute to our theory of mind, it does not yet have a fully satisfactory account of concepts. I am providing the causal and semantic components of concepts, there are other requirements that remain unfilled.

5.4.3.1 Pragmatic Conceptualism, or, Never the Same Concept Twice

I argued earlier\textsuperscript{610} that the activation signals are to be regarded as the locus of representation in the network, in that they can be regarded as being imbued with semantic content. I also argued that the set of nodes, weights and connections that make up the network can be regarded as comprising that network’s accumulated knowledge. In addition, a clarification as to how such knowledge might be regarded as explicit in a way that is distinct from the intuitive ‘textual’ notion was provided, and the processor itself was found to be a key role in identifying the degree of explicitness. Leveraging these resources, I now identify two different aspects of the concept ‘concept’ that bifurcate along mental and linguistic lines, which I shall call concept\textsubscript{mental} and concept\textsubscript{public}. Concepts\textsubscript{mental} shall be used to refer to the mental instantiations of concepts, and concepts\textsubscript{public} shall be used to refer to the publicly accessible semantic content associated with a concept. I now explicate these terms.

Barsalou has argued that the occurrent activation signals are not mere representations, but actually constitute concepts; concepts are “occurrent psychological states which drive categorization and typicality judgments … [which are] … constructed on the hoof. … the same concept is rarely, if

\textsuperscript{610} Clark, 1993, pg. 124

\textsuperscript{611} See Section 5.3.1
ever, constructed for a category”612. Barsalou is motivated to view occurrent signals as concepts in order to account for the ubiquity of typicality judgments in all sorts of ad hoc and theoretical domains. I take his point, and also want to view the occurrent signals as the proper locus for concepts\textsubscript{mental}, rather than mere representations. I am sympathetic to the idea that concepts\textsubscript{mental}, as they appear in the mind at any given time, are unique – both semantically and causally. Clark is loathe to take this position, however, as “this tendency to use ‘concept’ to indicate an instantaneous internal state (which can thus be implicated in the very simplest of causal stories) extends an unholy invitation to eliminativism.”613 That eliminativism is heresy614 I agree, and I take Clark’s worry seriously: if concepts are to be identified with occurrent signals, then there results a significant explanatory shortfall, namely the right to folk psychological explanation. One of the pillars of folk psychological explanation is that there exist mental items corresponding to the terms of behaviour explanation (a full analysis of the requirement of folk psychological explanation can be found in chapter 4). If eliminativism rears its head, the mental items which one looks to as a cause of behaviour are not to be found – no identifiable mental causation, no folk psychological explanation.

In my view, concepts\textsubscript{mental} can be seen as being instantiated at any one time by occurrent signals, in the sense that the signal is a manifestation of some concept\textsubscript{mental} at a particular point in time. So an occurrent signal is a concept\textsubscript{mental}, but a concept\textsubscript{mental} is not an occurrent signal in the same way that a carrot is a vegetable, but a vegetable is not a carrot. I do not want to deny that someone possesses a concept when one is not currently thinking about it, nor would I want to underestimate the role of background knowledge in concept formation and content. My proposal is to identify the set of all those occurrent signals that may counter-factually be produced by that network, given some appropriate input, as comprising the concept\textsubscript{mental} \ \cdot \ \text{Thus, concepts}_{\text{mental}} \ \textit{are some potential set of occurrent signals that are - or could be - produced by a given network}. There obviously remains a problem of concept individuation – which counter-factual set is associated with which concept\textsubscript{mental} – but I hold off on answering that question until I analyze concepts\textsubscript{public}.

Clark seems to endorse something along these lines, although his notion of concept does not delineate the mental and public aspects quite so deliberately; “… what we are doing when we ascribe to someone grasp of a concept is better divorced from considerations of particular occurrent states. Instead, we are crediting the person with a body of knowledge (stored, e.g., in the long-term weights of a number of subnetworks) which can power a variety (perhaps an open-ended one) of

---

612 Barsalou, 1987. pg. 114
613 Clark, 1993. pg. 93
614 I can’t resist: outside of Church(land).
occurrent states according to local factors.\textsuperscript{615} I have rendered explicit the otherwise implicit content of the weights and connections in allowing the set of occurrent signals to be counter-factually defined. Recall that information is to be regarded as explicit to the degree that a processor finds that information easily available. When a network is activated, the occurrent signals contain (a tiny portion of) that network’s implicit knowledge, to the extent that the information is now made available to the network – here the processor – itself. The counter-factual profile of the set of occurrent signals is meant to render explicit all implicitly present concepts\textsubscript{mental}. Horgan & Tienson appear to adopt a similar approach in advocating a ‘dispositional realization’ approach to what they call ‘cognitive forces’;

Higher-order structure is realized only in the overall counterfactual profile of potential temporal evolutions from one total lower-order state to another, a profile characterizable mathematically by a dynamical system. In the case of physical realization, abstract mathematical or cognitive structure is realized in the \textit{totality of potential physical processes of the physical system} – not in the intrinsic physical structure of physical states tokened or persists in the physical device ... Thus, ... cognitive ... structure is physically realized dispositionally. [emphasis mine]\textsuperscript{616}

What I am adding to this picture is making the activating signal an explicit part of how that counterfactual profile is defined.

What set of concepts\textsubscript{mental} one has is thus governed by two factors; the architecture of the network itself and the total set of signals presented to the network (past, present and potential). Note that as the architecture of the network changes, the set of potential occurrent signals also changes. I gain concepts\textsubscript{mental} over time as my network changes, and the form and shape of those concepts change concurrent with the network’s evolution. This should come as no surprise. A somewhat surprising result is the unabashedly \textit{externalist}\textsuperscript{617} nature of this account in that it is by virtue of being presented - or potentially being presented - with some particular signal that I am said to possess the

\textsuperscript{615} Clark, 1993, pg. 204
\textsuperscript{616} Horgan & Tienson, 1999, pg. 18
\textsuperscript{617} Note there are an additional two ways in which the set of concepts\textsubscript{mental} can be regarded as dependent – indirectly and directly - on the set of externally located stimulus signals. First, the set of weights and connections evolves according to the training set, and any occurrent signal is then shaped by those weights and connections; this point was brought out when I discussed strong representational change and the intermingling of code and process. The particular architecture instantiated by a network is highly dependent on the training set, both in terms of the items themselves and the order in which they were presented. Second, the occurrent signals that are later generated are directly dependent on the current stimulating signal, since it forms an integral part of the context that shapes the content of the signal.
associated concept\textsubscript{\textit{mental}}. Intuitively, this can be regarded as opening the door to public semantic content. Note though, that it cannot be said that I did not possess that concept prior to the presentation of the signal, because the occurrent signal generated was always a member of that counter-factually designated set, even though the stimulus signal had not yet been tokened.

A serious problem arises here in delineating existing (known) concepts and new (learned) concepts. The problem can be put thus: how might it be that new signals, representing new concepts, cannot be said to already exist in the counter-factually defined set of concepts\textsubscript{\textit{mental}}? If potential signals are to be included, how am I to say that I do not already possess all of the concepts\textsubscript{\textit{mental}} that I could ever have in my possession? Does this theory not fall prey to the same massive nativism that plagues Fodor’s account?

The problem here relates to \textit{concept acquisition}, and what it means for a network to be able to exhibit unsupervised learning. My response is to argue that all novel signals are – until a concept is learned – categorized into existing concepts; that is, they are transformed according to some existing topologically described dynamic structure. New stimuli are regarded to be manifestations of a variant of old stimuli; since we perceive nothing unconceptualized, new objects initially fall into old conceptual categories. We see new objects as previously existing types of objects – until, of course, we have learned to categorize the new object in a novel way. Until that occurs, I want to say that the counter-factual profile is dominated by network architecture – the person \textit{cannot} manifest an occurrent signal that would correspond to the new concept. Hence, the counter-factual profile does not already contain the new concept. Massive nativism is avoided. The problem has been put off to the unsupervised learning phenomenon. I have already stated that C-Nets are capable of such a feat, and here I would simply add that such a feat alters the topologically described dynamic structure of the network, such that new occurrent signals – reflecting the new conceptual category – \textit{can} be generated. The counter-factual profile of the set of occurrent signals is thus updated.

Recall Clark’s worry of eliminativism, and the potential loss of folk psychological explanation on a ‘concept equals occurrent signal’ view. The problem centers on whether or not such a signal can be accorded some modular, propositional content to fill the causal role posited by folk psychological theory. In its most blunt form: when I get a carrot from the refrigerator, the thought that is causing my behaviour – on my account the concept\textsubscript{\textit{mental}} corresponding to ‘carrot’ – had better be about carrots (and not dogs), and had better be what causes the behaviour. The question relates to the sort of relationship that holds between the counter-factually defined set of occurrent signals and the semantic content itself, and on what basis the set of signals can be said to exhibit any sort of \textit{unity}. If the causal account is a chaotically disjunctive set, so the worry goes, there is no
good reason to think that the causal role attaches to the folk-psychological picture – on this account, if concepts_{\text{mental}} are causal (but chaotically disjunctive), then they are simply not the same thing we are talking about when we talk about concepts. Neurological co-relates of carrots are not what we talking about when we speak of ‘carrots’, and the philosopher of mind has a duty to take a stand on this issue.

I do not think that the validity of folk psychology depends - whatsoever - on whether or not our best neurological story will find the requisite co-relates. Indeed, on the view I am advocating, it will not find such co-relates since every instantiation of concept_{\text{mental}} is in all likelihood both temporally and inter-personally unique\(^6\). I’m perfectly happy to admit that every time I go to get a carrot, or exhibit any other behaviour in folk psychological terms – including linguistic behaviour – that the causal story told at C-Net level will be unique. It might be the case that a story that is overtly causal in nature may not be particularly explanatory; these are not, however, mutually exclusive categories, but categories that are merely somewhat unhappy bedfellows. The level of causation is simply the wrong place to look for an explanation along the folks’ lines, but with some additional resources we may see how they might be linked.

I now differentiate concept_{\text{mental}} from concept_{\text{public}} and argue that folk psychology might be saved if I can meet two explanatory demands: first, that our respective possession of concept_{\text{public}} (and not our concepts_{\text{mental}}) be sufficiently similar (not identical\(^6\)\(^9\)\() across persons and times to account for folk explanations of behaviour, and second, that there be some philosophically\(^6\)\(^2\) respectable way of building a relationship between concept_{\text{public}} and concept_{\text{mental}}. It simply does not matter how different concepts_{\text{mental}} vary across persons or across time in the same person if these two conditions are met.

The first explanatory demand is simple (in one respect): concepts_{\text{public}} are just defined as those semantic items that we share, which we identify by the use of language (that’s the easy part). What those semantic items refer to - from an ontological perspective - may come in any variety of flavours (that’s the hard part); if one is medieval in one’s thinking, perhaps they correspond to natural properties, if one is idealistic then perhaps they correspond to Platonic Forms, if one is

\(^6\)\(^1\) There is current evidence (see Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C. & Fried, 2005) that singular neurons firing can be associated with singular semantic events – data that would seem to undermine this claim. However, there is no reason to think - no matter how sparse the number of neurons associated with some semantic event, take the ‘grandmother neuron’ as the end-game of this account – that there is not all sorts of contextual neural firing, prior during or after, that must also take place for that sparse event to play the functional role that it does. A single neuron may fire on a given semantic event, but what else must transpire for that sparse event to be an event of the mind?

\(^6\)\(^9\) That concepts_{\text{public}} be merely similar across persons, and not identical, I shall address presently.
inclined to think nature carves nicely at the joints then perhaps they correspond to natural kinds. I have argued in Chapter 2 for a view of ontology such that our concepts\textsubscript{public} correspond to the patterns that are the variously identified swirling soups of Ontology I, the identification of which is useful and suited to us, our purposes and our capacities, but much of my argument about concepts and intuition can be separated from that claim.

I take it that concepts\textsubscript{public} are imbued with semantic content by whatever pragmatic and empirical means we have at our mutual disposal; be those means referrals to experts (a ‘quark’ is what our best physicists \textit{tell} us is a quark), gesticulations of a mother to child (‘\textit{that’s} what I mean by ‘table’), or the open exchange of ideas in the academic, political or larger social dimensions. Recall that a normative aspect to pattern-recognition - operationally defined as whatever a pattern-recognizing creature may recognize as a pattern - was captured by viewing that creature as being embedded in a socio-linguistic community. Hence, it is our \textit{shared public life} – of which language is a component – that ensures that concepts\textsubscript{public} are shared amongst members of a sociolinguistic group. This is the ‘pragmatic’ portion of pragmatic conceptualism. The means by which the semantic content of concepts\textsubscript{public} is sharpened is by public dialogue and activity, and the degree to which they can be made precise is dictated by the nature of the concept – it may be the case that mathematical concepts can made much more precise than political concepts, for example.

There remains an important issue to resolve. I must somehow account for the possibility that I can be mistaken in what I hold a concept to mean, or that I may have some incomplete notion of it. One person’s concept is not the same as another’s. My concept of an electron is not the same as the physicist’s even though, on my account thus far, we both seem to share the same concept of ‘\textit{electron}\textsubscript{public}’ by virtue of sharing the same \textit{word}. My concept of statue may not be the same as my friend Chris’, although we seem to share what I have identified as the concept \textit{statue}\textsubscript{public}. Concept\textsubscript{public} meets what Fodor called the \textit{publicity} requirement, but I accused Fodor of overstating the degree to which our concepts are shared – it seems I might be guilty of the same crime.

It is here that we see why the possession of a concept is likened to the possession of a skill. Skills may be possessed to a larger or lesser degree, and the standard-bearer of a skill – that which stands as the objective standard of excellence - may not actually rest within any one person (although it might!) but rests within the community at large. Skills are pragmatically defined relative to some task or objective, and their possession is \textit{demonstrated}. Concepts\textsubscript{public} stands as the objective, standard-bearer of semantic content and any individual possesses that concept\textsubscript{public} to the extent that they are capable of demonstrating that possession. Any individual partakes of the concept\textsubscript{public} to a

\textsuperscript{620} Contra - scientifically.
larger or lesser degree. At a bare minimum, being able to say the word when it pops into your mind enables one to begin demonstrating and honing the skill. Saying the word at the wrong time, or in inappropriate circumstances, indicates an unskillful use of the concept – and the sociolinguistic community would respond by applying the appropriate sanctions, approbations, corrective measures, etc. Being able to provide a definition, properly identifying the appropriate patterns associated with the concept, teasing out repercussions and implications of the concept – these are all manifestations of the skill associated with its possession.

It is easy to say one possesses a portion of a concept, or that one’s concept is incomplete or distorted. These ways of speaking all correspond to the view that possessing a concept is like possessing a skill, although what people actually possess - physically possess - on my view, is some manifested portion of the counterfactually-defined set of occurrent signals that is concept mental. The possession of a concept is enabled by virtue of successful participation in a life shared with other members of a socio-linguistic community. We also want to be able to say that one’s concept can be mistaken - although this way of speaking begins to stretch the metaphor of skill-possession. Can a skill be mistaken? No – but one can bring to bear an irrelevant skill, or the wrong skill at the wrong time, and so it seems harmless to keep this terminology.

Where does the concept reside, and what is it? This is a complicated question, and I can only point to an answer. Our books, culture, artifacts, activities, institutions, language, conversations – these all form part and parcel of the answer, and these all contribute to the excitatory signals that spur manifestations of our respective concepts. It is in our shared culture that our concepts reside. Correct concept possession accords with successful interactions with the socio-linguistic community at large.

There remains an objection, one for which I can only point to a solution. If concept is to be thought of as that which is possessed by a socio-linguistic community and reflected in their behaviour and culture, then it is entirely possible that a concept might be simply wrong. Content resides not in some fact-of-the-matter, but what some community believes to be the fact-of-the-matter. This objection has some force, and must be taken seriously. I dodge this bullet by adopting a suitably pragmatist position, and posit that the community-at-large is engaged in some on-going process of getting ever closer to the truth. A socio-linguistic community will, following Peirce, use a number of salutary methods of investigating the world around them, and shall not remain under collective illusion or with incomplete evidence for long. Humans are constrained, in suitably hard
measure, by the world around them and cannot long maintain conceptual error – assuming, that is, 
that they adopt those salutary methods of investigation!

The second explanatory demand – that of linking concepts\textsubscript{mental} to concepts\textsubscript{public} - I answer by 
proposing that when those words appear as concepts\textsubscript{mental} within our mental economy - in the form 
of some internal dialogue that marches through our minds on any given day, for example - those 
concepts\textsubscript{mental} may be regarded as inheriting their content from public exchanges. I think this move 
is made possible by the fact that I am capable of bringing into awareness the overt and public 
nature of their origin. My own mental dialogue, consisting of words which act as labels for 
concepts\textsubscript{mental}, is unique in that each occurrent signal is unique, but it not unique in terms of how I 
choose to interpret that signal. The signal appears to me as having semantic content, it takes the 
form of a thought - it does not appear to me as some unique pattern of energy traveling through my 
network (what would that be like?).

However it is accomplished, it is an empirical fact that my brain is capable of interpreting at least 
some of its own activity, or some of the occurrent signals passing through it. How this may happen 
I do not know, but on a network-of-networks account, C-Net resources may be capable of 
accounting for an interpretation of inter-network signals as identifiable words or labels, just as a 
more locally-defined network is capable of interpreting some signal as an object. This is what (at 
least one aspect of) consciousness amounts to, for me - it is a kind of self-reflective awareness of 
our own mental activity, and it generates a perspective on our own internal thoughts that is 
reflective of the community at large. It is reflective of the community in the sense that one interprets 
thought from a socio-linguistic perspective, the words that came from the public arena echo in my 
head just as the images that flash in my mind have their origins in my perception of the external 
world. Thus, at least one role that consciousness plays in our mental lives is to bridge the 
self/socio-linguistic-community gap by providing (awareness of) a semantic interpretation of 
concepts\textsubscript{mental}.

So – an internal signal is presented\textsuperscript{622} to me (somehow!) as a semantic item – a word - and it is the 
fact that the signal appears to me as a word which allows each of the multitude of concepts\textsubscript{mental} 

\textsuperscript{621} What portion of some concept\textsubscript{public}, does an individual possess? That is a complicated question that can only be 
answered in the on-going dialogue that is human interaction. Perhaps that’s exactly what university exams try to 
measure.

\textsuperscript{622} I have self-consciously used the word ‘presented’ here, but I do not mean to imply the existence of that old 
humbug homunculous to whom something is presented. Because I am offering no explanatory mechanism for the 
way in which consciousness is playing the functional role to which it has been assigned, I almost want to say 
something like ‘I participate in the signal’ whereby the fact that signals appear as contentful is analogous to a sort of 
developed cognitive skill.
that fall under the same label to share that same content. I am capable, if pressed, to explain what a
word means; I have at my disposal a set of knowledge about that word. That knowledge is derived
from past associations with that word that came to me in the form of signals from the external
world, moderated by my linguistic community, it resides in a set of weights and connections, and it
can be expressed by the generation of some large, potential set of relevant occurrent signals. Since
countfactually defined set, I do not need to actually bring all of my past,
public encounters with the word into consciousness, but the fact that I could means that those
encounters are already a part of the set of relevant associated with that label. Hence, I
claim that my ability to interpret my own signals as the words that we publicly share, and potentially
access the causal-historical path by which I have shared that word in that public life, provides a
philosophically respectable means to build the required relationship between concepts
mental and concepts
public.

Consciousness is obviously playing a key role here, and I am not pretending to solve the
consciousness riddle; quite the opposite – I am leveraging consciousness to solve the problem of
semantic/causal mental content. I am really only saying that it is by virtue of having concepts
mental play out their lives on the stage of consciousness, and thus having the (magically unexplained)
ability of consciousness available to provide the necessary semantic interpretation, that the link
between the public and private roles of concepts may be bridged. Note though, that the bridge is
built only relative to a single processor; one is capable of interpreting only one’s own mental states.
If you wanted to know what content my signal has, it could only be made explicit by activating my
system, by putting the right processor into the picture. This is accomplished not by examining my
C-Net yourself – nothing is made relevantly explicit from that perspective– but by watching me
behave. Give me some input, watch my output, and extrapolate what content my internal signaling
may have. The observed behaviour may be linguistic, and the input to my network may take the
form of a question; I may simply tell you what is causing my behaviour when you may ask. I might
lie to you, or be mistaken, but that does not matter to the question at hand: the relationship between
concept
mental and concept
public for anyone other than oneself is shaped by some open-ended
interpretation of public behaviour.

The locus of causation is to be found in the occurrent signals, and that claim I take as relatively
unproblematic. Here, I’ve outlined a way in which the folk psychological picture touches this causal
picture, but the two pictures do indeed bifurcate along the concept
mental and concept
public lines; the
folk psychological content may have a well-defined locus of activity (concept
mental), but in order to
fill in what precise content that locus of activity carries at any one time, one is required to raise the
view to the level of the activity of the network. A person can do that to themselves, but a person
cannot do that to another.

In other words one can claim that a folk explanation is causal, and that the locus of the activity
which contains that folk content may be found in the occurrent signal, but it doesn’t work the other
way around. One cannot get to the folk level of content by examining some particular signal that
would correspond to the locus of activity associated with the causal component of that content. On
this view you can move from content to causality, but not from causality to content. Clark; there
may be “no unity [among the occurrent mental states associated with some concept] visible without
the lens of folk-psychological interests”. Identifying the concept \( \text{concept}_{\text{mental}} \) as playing a causal role allows
me to deny that I am a behaviourist, but it is only in overt behaviour that may we regard the causal
story meeting our psychological explanatory needs.

How the relationship between concept \( \text{concept}_{\text{mental}} \) and concept \( \text{concept}_{\text{public}} \) may be made more precise, I have no
idea. That there may exist the famous “grandmother neuron” is an empirical question, and I take it
that there will be great strides in characterizing the relationship between contents of thought and
their neurological co-relates. If I am right in my view of C-Nets though, there will remain a very
loose relationship between the semantic item and the possible set of underlying instantiations
defined from a neurological perspective, particularly as the set of instantiations may vary from
person to person. How much they would overlap between persons would be a function of the
similarity of the associated training set, among other things (like initial conditions of the network,
for example). Our concepts \( \text{concept}_{\text{mental}} \) of Elvis Presley or Brad Pitt may be much more similar than our
concepts \( \text{concept}_{\text{mental}} \) of ‘justice’ or ‘shrub’, since our encounters with the former may be quite similar
(how many different pictures of Elvis or Brad are shown on the evening news?) while our
encounters with the latter may be quite different.

5.4.3.2 Unsatisfied/Further Explanatory Obligations

It is a broadly understood drawback of C-Net accounts that a general theory of concepts is
currently unavailable if that theory is understood to encompass the whole gamut of explanatory
demands generally associated with the publicity, systematicity, productivity and causal-role
requirements so thoroughly explained by an LOT account. My account of semantic and causal
roles given above notwithstanding, it is a pretty universal demand that concepts are discrete and
recombainable – and all of this at the level of the items defined in our folk psychology. This is the
most serious direct cost the C-Net theorist must pay for losing the ‘classical cascade’. Unless the
connectionist finds a way to give a “general account of conceptual combination [which] involves the systematic recombination of grasped meanings”\textsuperscript{623}, then the account remains incomplete. It is one thing to claim that C-Nets can account for intuitive judgments about concepts, and to argue that those judgments may involve the use of theoretical terms, but “the ability to entertain any one of an infinity of potential thoughts involving the systematic recombination of grasped meanings is another. … No connectionist has, I believe, a convincing story to tell about these matters.”\textsuperscript{624}

There are, of course, some equally well-understood drawbacks to the classical account itself. Aside from the arguments I have made about the nature of the (non)computability of patterns\textsuperscript{625}, and the overly strong characterization of the systematicity and productivity requirements\textsuperscript{626}, there is also the problem of relevance as it relates to the question of how a system might be made sensitive to global parameters\textsuperscript{627}: Clark; “How is the right complex thought generated at the right time?”\textsuperscript{628}. Belief formation is subject to properties of the entire belief system (as we have learned from Quine), and we have seen that any part of a person’s knowledge may be relevant to some current interpretation task. This remains a problem for the classical account. Clark;

... globally sensitive processing runs classical systems very quickly into well-known problems of combinatorial explosion … it is simply not clear that the classical approach actually solves the full-fledged problem of the productivity of thought. It gives a neat account of the structural relations between (putative) vehicles of thought. But the price of that account seems to be an equally deep problem concerning actual on-line production: It is no longer clear how the right thought complex is produced at the right time.\textsuperscript{629}

Horgan & Tienson (H&T) take a similar line in declaring:

The potential relevance of anything to anything, that is, the potential relevance, in some context or other, of any of the unlimited number of things a human being can represent to any other thing the human can represent. The fact that both scientific confirmation and individual belief fixation are determined in part by such factors as simplicity, conservatism, and plausibility, factors that depend upon global properties

\textsuperscript{623}Clark, 1993, pg. 112
\textsuperscript{624}Clark, 1993, pg. 108
\textsuperscript{625}See Chapters 2 and 3
\textsuperscript{626}See Chapter 4
\textsuperscript{627}See Chapter 3
\textsuperscript{628}Clark, 1993, pg. 109
\textsuperscript{629}Clark, 1993, pp. 109, 111
of the current belief system. No one has any idea how cognition with any of these three characteristics might be simulated using programmable representation-level rules.  

I take it that these failures of the classical account diminish, to some extent, the demands on the connectionist to come up with a theory of concepts that can account for the varied characteristics of our higher thought, since that competing theory has yet to meet its own explanatory obligations. This, of course, is merely a dialectic move involving a burden of proof issue however, and should not detract us from seeing that deep problems remain for the C-Net theories.

The problem remaining for the C-Net theorist has mainly to do with needing to account for the structure of our thinking, in identifying concepts as having identifiable, discrete and redeployable content; such content that can be made generally accessible for recombination with other semantic items. I have indicated how such semantic items might be generated, and how they might be combined in a “single, preexisting representational space”, but until this account can be extended so as to be made more general, it must be seen as unsatisfactory. I have four responses to this unhappy situation, which correspond to the following ideas: first of all, a diminishment of the demands of structurality, followed by three ways of buttressing the structural resources available to a C-Net theorist; a generalization to a network-of-networks model, leveraging the role of consciousness, and a greater reliance on language as an externally-derived cognitive tool. All of these responses are tentative and hand-waving in nature - they are here to indicate where I would want to further explore this sort of thinking.

First of all, I have argued above that the degree of systematicity that need be accounted for is less than that which Fodor (and by extension other classicists) demand – and which I have characterized as demanding a crystalline structure for thought. Lowering the bar to mere structured thought (recall, more like a tree than a crystal) increases the likelihood of a C-Net account taking root. Clark has pointed out that the activation forces in a C-Net are not unstructured, but have a structure that corresponds to a semantic network. Horgan & Tienson (H&T) generalize this characterization to a “cognitive forces” model and argue that “the basis of cognitive processing is content-appropriate cognitive forces”. For H&T this is an appropriate level of structure because they argue that “the important theoretical generalizations of intentional psychology are or should be

630 Horgan & Tienson, 1999, pg. 12
631 Clark, 1996, pg. 112
632 See Section 4.4.1
633 Horgan & Tienson, 1999, pg. 12
ceteris paribus generalizations [which] constitute a distinct logical form.”

All intentional explanations permit exceptions (depending on context) and have at their heart a kind of generality that is not reflective of the sort of hard-and-fast rules-based approach associated with the LOT story. Rather, the generality is reflective of the “basic causal tendencies of cognitive systems [which] come in systematic patterns.”

The upshot of all this is that the structure of intentional explanation is best characterized by what H&T call “soft laws”, and this is because the underlying cognitive system is connectionist in nature.

Having lowered the bar over which the C-Net theorist must hop (in terms of the degree of structurality that must be accounted for), I now point out three ways in which the network-level structurality may be supplemented. The idea here is that - combined with the lowered explanatory demands – these supplementary ways of generating cognitive structure will be sufficient to allow for a full-blown C-Net theory of concepts to go through.

I am quite happy to hand-wave at a network-of-networks model of the mind in order to make more general a solution that seems close at hand. I note again that Clark admits that the connectionist succeeds in this systematic, generalized concept recombination “when the combined content lies within a single, preexisting representational space” and he has availed himself of theoretical inputs traversing these networks. All I’m really saying is let there be many networks, sharing signals among each other, and what Clark regards as a local solution may generalize. Occurrent signals have been imbued with semantic content, and in being made available as cross-network traffic are free to combine with any number of other signals on the usual ‘cognitive forces’ model. Further work needs to be done to explain how one signal, emanating from one sub-network (with its attendant semantic metric) may be combined with another signal, emanating from its (presumably distinct semantic) sub-network, but some basis for concept recombination has been established. Clark is much more reticent about these sorts of moves, but I have stated that these are exploratory moves in which I want to stretch out the C-Net resources a bit.

Second, I would posit that – just as I attempted to leverage the use of consciousness in bridging the gap between concepts_{mental} and concepts_{public} – consciousness may be able to play a role in accounting for the demands of structure in our thought. We are perfectly capable of judging what sorts of concepts go with others in a structured, ordered sort of way, just as we are equally capable of thinking nonsense thoughts (‘orange I predicting floofa eggs today monkey in’). We don’t often think in unstructured, nonsense sorts of ways - it is neither interesting nor useful – but it

---

634 Clark, 1993, pg. 14
635 Clark, 1993, pg. 14
seems to be a cognitive possibility that thoughts may be unstructured as much as they are structured. It does not seem to me that my mind is automatically quite as structured as the classicists would assert, rather, it needs to be taught to be structured, both by a series of teachers as well as by a rigorous attempt by myself to properly negotiate whatever conceptual terrain I need to negotiate. In each of these cases, the mind is shaped by conscious, internal judgments of relevance, coherence, etc. and in that sense, consciousness (both mine and my teachers’) may be seen to play a role in shaping what sorts of thoughts are to be associated with what other sorts of thoughts. In this way the structure of thought may emerge as the result of a meta-network view, enabled by consciousness, of passing internal signals. Those signals need to fit certain patterns, and so consciousness plays the role of building the correct pattern-recognition responses at a level of inter-mind communication, or at the level of internal dialogue. My internal dialogue seems to reflect – roughly – the dialogue that I hear all around me, and so perhaps this is just another case of pattern-recognition. Roughly, the internal flow of concepts is regulated from a pattern-recognition perspective relative to dialogue around me, and hence makes use of the same sorts of computing resources as the lower-level networks. Language, of course, plays a central role here, since that is the medium by which patterns of thought are consciously judged to cohere to the external scaffolding that is my socio-linguistic community.

So - third, and probably most promising - is to posit that the structure of our thinking is inherited - and reflects – the structure of public language, rather than vice versa. This is a view shared to different extents by Place, Dennett and Clark (among others). The basic idea is that it is language that shapes the operation of our minds, rather than being a reflection of those operations. Fodor clearly espouses the opposite view; for him, the systematicity of language reflects, and is caused by, the systematicity of thought. So too with Chomsky, for whom the emergence of language in a child is reflective of that child gaining access to a generative grammar that is stored

---

636 Clark, 1993, pg. 112
637 Place; “Not only is there reason to think that there is no role for either syntax or logic before linguistic competence is acquired; there is reason to think that when syntax and logic eventually appear, they are learned by the child, not as internal principles which govern the generation of thought and language, but as external constraints to which the brain must learn to conform ... in order to secure the appropriate reinforcement from the listener in the form of an acknowledgment of successful communication.”, Place, 1994, pg. 76
638 Dennett seems to view language as an external source that modifies the brain’s structure (presumably as a kind of massive training set) so that the emergence of conscious, rational thought is something like “serial virtual machines implemented-inefficiently-on the parallel hardware that evolution has provided for us”, Dennett, 1991, pg. 278
639 Clark views language as “in essence just a tool – an external resource that complements but does not profoundly alter the brain’s own basic modes of representation and computation”, Clark, 1997, pg. 198. In speaking as the brain of John – a fictional fellow with a body, who inhabits our socio-linguistic world, Clark muses “Possessed as John is of such a magnificent vehicle for the compact and communicable expression and manipulation of knowledge, he often mistakes the forms and conventions of that linguistic vehicle for the structure of neural activity itself.”, Clark, 1997, pg. 227
deep within the language facility. Perhaps the truth is somewhere in the middle, and the more interesting question would have to do with the sort of pre-conditions that must exist to allow for language to take hold. Clark;

If systematicity falls out of the use of structured language, the question of the preconditions for the robust emergence of systematic thought reduces to the question of the preconditions for the robust emergence of linguistic (especially systematic) knowledge. And minimal nativist conditions ... for the emergence of knowledge of syntax will typically fall far short of positing anything like a language of thought.\textsuperscript{640}

On this alternative view, at least some of the systematicity that is to be found in our brains is best viewed as an emergent phenomenon, dependent on exposure to language, which can itself be regarded as a kind of external tool that – over time with training – becomes internalized (either as a tool, following Clark, or as a source of computational modification, following Dennett). Obviously, this is not an explanation – one would need to explain how exposure to language generates a parallel kind of systematicity in the brain – but it is certainly a promising method of alleviating the explanatory burden on the C-Net theorist by providing her with a source of the systematicity that is to be explained, rather than forcing her to explain how that systematicity emerges \textit{de novo}.

\textbf{5.4.4 Occurrent Signals as Intuitive Judgments}

Recall that Barsalou advocates viewing the occurrent signals themselves as concepts. I have argued that this move will not suffice, since – among other reasons – one presumably has a concept when it is not an occurrent thought. The point of Barsalou’s proposal though, was to account for the prevalence of typicality judgments. There is something right about Barsalou’s position, in my view, and I would be happy to equate my notion of intuitive judgments with his notion of concept (as the source of typicality judgments). Concepts may be fuller, more fleshed out things than these mere category judgments (is judging membership really all we do with our concepts?) but the same cannot be said of what I am calling intuition. Recall that intuition (O-Intuition or H-Intuition) is defined as the categorization of some phenomena as a pattern of a determinate type, the categorization of which is propositional in form (thus attached to a linguistic label) and which appears as epistemically necessary (it does not appear the intuition could have been otherwise) and

\textsuperscript{640} Clark, 1993, pg. 225
unsupported by an inferential process. Note that I am not claiming an intuition is equivalent to a
typicality judgment; rather, I am claiming that the spirit of Barsalou’s proposal seems a good
starting point for my characterization of intuition, which I take to be a broader notion than typicality
judgments, but which shares with them a kind of phenomenological immediacy and pattern-
recognition character. Thus, my proposal is to equate intuition with (a relevantly-defined subset of)
the occurrent signals in the network. There are two qualifying remarks about what sub-set of
occurrent signals is relevant.

Before those qualifying remarks, however, I do want to make very clear that I am not defending any
sort of type-type identity theory in equating intuition with that relevantly-defined subset of
occurrent signals. Indeed, my position is quite the reverse: because of the nature of the set of
signals I am arguing against any sort of type-type identification. I made it very clear in the
previous section that the counter-factually defined set of occurrent signals that comprise some
semantic item (the type) is a chaotically disjunctive set, and so cannot be identified by any means
other than the introspective, conscious judgment of the thinker. There is simply no way to move
from the level of the signals to the level of the semantic item, other than to be a living creature
experiencing those thoughts as of a discrete semantic item from occasion to occasion. This is part
and parcel of the nature of a chaotically disjunctive set; the defining conditions of the set are not
identifiable at the level of description of the signals themselves. In other words, there is no way to
define the boundaries of that set, other than by conscious experience. Hence, not only do I not
endorse a type-type identity theory, but I argue that such a theory is false if connectionism is to be
taken seriously.

One defining mark of the occurrent signals of interest, is the emergence of these signals into
consciousness. Intuition is - by definition - a conscious cognitive act. Not all occurrent signals will
be conscious, but those of interest will be. How that happens - what set of necessary and sufficient
conditions must obtain (what sort of functional role the occurrent signal must play within the C-Net
environment? What sort of dynamic characteristics must obtain over that signal?) I cannot say in
any detail - but one thing does seem clear: only signals with determinate semantic content are
eligible candidates for appearance within consciousness. That is, we are always conscious of
something. Hence, I differentiate between the two conditions for an occurrent signal to become a
conscious item: call these conditions the conscious necessity and conscious sufficiency conditions. I
shall provide the necessity condition but shall not provide the sufficiency condition.
We can only be conscious of the output of some network (as some appropriate categorization of the input) since the only signals in the network that are endowed with determinate semantic content are those input signals that have been transformed by a network and appear as an ‘output’, appropriately labeled. This is the case because of the indeterminate semantic character of the individual nodes themselves on the super-positional storage model\textsuperscript{641}. Expanding on this idea: \textit{occurrent signals composed of previously transformed signals are the only signals that may qualify for consciousness – previously transformed signals consist of two types: inter-network signaling (on the network-of-networks model) or intra-network signals on a highly recurrent model (in which outputs are stored as context-layers for later presentation to the input).} So the first qualifying property of the subset of occurrent signals of interest is now both phenomenologically defined (as being conscious), and functionally defined (as limited to those transformed signals that play the role of an output). It is thus a necessary – but not sufficient - condition for consciousness that signals be appropriately transformed. Call these signals \textit{transformed occurrent signals}.

A second defining mark of the occurrent signals of interest is that they are to be signals with content for which the conscious ‘owner’ of the network has no direct introspective evidence or what I previously called \textit{cognitive support}. Intuitions are conscious thoughts, but not all conscious thoughts are intuitions – some thoughts do not appear bare and isolated on the stage of consciousness, rather, they are part and parcel of some inferential train. It was the \textit{phenomenological} character of the intuition that made the difference when intuition was defined in Chapter 1, and we are now in a position to flesh out that distinction – albeit in a very general way - in C-Net terms. Recall that there are four possibilities with respect to the underlying cognitive processes which produce a conscious judgment:

\begin{itemize}
  \item[a)] There is an underlying inferential process, but no conscious access to it.
  \item[b)] There is no underlying inferential process, and no conscious appearance of one.
  \item[c)] There is an underlying inferential process, and conscious access to it.
  \item[d)] There is no underlying inferential process, but there appears to be access to one.
\end{itemize}

In cases a), c) and d) there is cognitive support, and in case b) there is none. In the case of d), it seems that the apparent inferential process is a post-hoc rationalization or a \textit{generated fiction} of the network. The presence of an underlying inferential process can now be made explicit: only those \textit{transformed occurrent signals} composed of other \textit{transformed occurrent signals} qualify as having

\textsuperscript{641} See Section 5.3.1
underlying inferential processes. On this view the *reason* there is no evidence available to introspection as to *why* the intuition takes the form that it does, is because the features and properties that contribute to the virtual prototype by which membership is graded (for example) are not even defined as independent semantic items available for appraisal. More generally, there are no context-independent, informationally discrete items to play inferential roles. The question as to which inferential processes do appear to consciousness, and which remain hidden, I cannot answer—this is the *conscious sufficiency* condition I referred to, but could not identify, above. I now apply the *conscious necessity* condition to whatever occurrent signals are causally responsible for the production of the occurrent signal of interest. Hence, transformed occurrent signals that consist in other transformed occurrent signals only qualify as intuitions if the *conscious sufficiency condition* is not met. Transformed occurrent signals that do not consist in other transformed occurrent signals always qualify as intuition. Hence, intuitive judgments are those occurrent signals that meet the necessity requirement (are transformed signals) and which are either i) not composed of other signals that meet the conscious necessity requirement or ii) composed of other signals that do meet the conscious necessity requirement but do not meet the conscious sufficiency requirement. Call his subset of occurrent signals *intuitive* occurrent signals.

Hence, intuitive judgments are those occurrent signals that meet the necessity requirement (are transformed signals) and which are either i) not composed of other signals that meet the conscious necessity requirement or ii) composed of other signals that do meet the conscious necessity requirement but do not meet the conscious sufficiency requirement. Call his subset of occurrent signals *intuitive* occurrent signals.

Like concepts, intuition bifurcates along public and private lines. Since intuitions are—by definition—spontaneous judgments that are *propositional* in character, the semantic content of the intuition may be expressed as intuition\textsubscript{public} and the occurrent signal itself may be denoted as intuition\textsubscript{mental}. Intuition\textsubscript{public} may be regarded as inheriting the public semantic content of the linguistic label; again, it is our capacity to be *aware* of the content of the intuition that enables this bifurcation to be bridged. Note that intuition\textsubscript{mental} is a *single* episodic element of an occurrent signal, and not a counter-factually defined set of such signals and as such, it cannot be assumed that the intuiter may possess intuitions prior to experiencing (or generating) them (*contra* concepts).

Since the occurrent signal that is intuition (intuition\textsubscript{mental}) shares public content with its conceptual counter-part (concept\textsubscript{mental})—this is true by definition, the propositional content of the intuition ensures it is captured by a concept\textsubscript{public}—it can be seen that intuition\textsubscript{mental} is a member of the counter-factually defined set that is one’s concept\textsubscript{mental}. Intuition\textsubscript{public}, then, may be regarded as direct evidence of the concept\textsubscript{public} one holds, and hence cannot be mistaken. One’s intuitions are indeed expressions of one’s concepts. This result should not be surprising: intuition\textsubscript{mental} is the result of an external ‘prodding’, specifically *designed* to reveal one’s concept. When we construct Gettier-
cases, we are – precisely – soliciting the concept held by the person being queried. We are presenting a signal, in the form of the imagined case, and soliciting a response or classification by the network, in the form of a responsive signal.

In this sense, a more sympathetic reading of the IDR may be noted: our intuitions do indeed provide evidence of the content of our minds – of our concepts - however, that content is only imbued with semantic content by association with the public component of concepts. It is not to the mind that we look for the source of semantic content, it is to the public sphere so – on my view – IDR has it exactly backward: concepts_{mental} (evidenced by the appearance of an intuition_{mental}) inherit their content from concepts_{public}, not vice versa\(^6\).

Intuition_{public} may be infallible as evidence of one’s concept_{public}, but one’s intuition can be mistaken just as one’s possession of a concept can be incomplete (or mistaken – recall the skill metaphor allowed us to retain the terminology mistaken). One can misapply a concept_{public} and so one can misapply intuition_{public} – one can be, simply put, wrong. Presumably, when I make an intuitive judgment about some state of affairs, we want to be able to say that there is a fact of the matter about that state of affairs – it is or isn’t an instance of knowledge, there is or is not a weak king side facing me, etc. While it is true that there are ambiguous situations about which we would want to withhold judgment (is that, or is that not, a threatening sky?), I take it the reason these situations are ambiguous is because there would be no consensus, and hence no good answer.

These judgments are context-dependent in precisely the same ways in which the signals themselves are context-dependent. The context in which their content is to be transformed into some category – the context in which the pattern is to be recognized – includes the surrounding context in which the item of interest is embedded within the original input signal, as well as other signals within the network (on the recurrent network or network-of-networks model). Note that these sorts of contextual items may – or may not – be contenders for conscious purview. In other words, much of the context that one takes into account informing an intuitive judgment may be unconscious (and arbitrarily subtle)\(^6\) cues. In addition, it can be seen that a part of that context must include all past encounters with similar patterns, since it is that causal-historical pathway that has determined the set of weights and nodes (the knowledge of the system) into which it has settled. This last sort of

\(^6\) There may be exceptions to this rule, of course, such as a case in which the concept is some particularly phenomenological type, such as ‘the feeling of dread’; such a concept is rooted in phenomenology and I would certainly accept that the content of such a concepts owes its derivation to a mentalistic source.
context-dependence is a direct result of the code-process intermingling so defining of the distinctive character of C-Net processing.

643 Recall the importance and repercussions of non-linearity in these systems. C-Nets may be tuned to be arbitrarily sensitive to aspects of input signals, and there is no guarantee that the cues to which it has tuned have ever crossed the consciousness threshold by fulfilling the conscious sufficiency requirement.
Chapter Six

6. INTUITION AS EVIDENCE REVISITED

Recall the original question to which this project is addressed: what is the proper role of intuition in philosophical analysis? I have limited the purview of intuition to be that of spontaneous judgements of fact, propositional in form, for which introspective evidence is unavailable. I remind the reader of the final definition of intuition, given in Chapter 1:

**Intuition:** direct, immediate, spontaneous and reliable mental judgments ‘that p’, in propositional form, that appear epistemically necessary; that is, they appear as if they could not have been otherwise.

I have argued that this definition – phenomenological in character - encompasses a number of cognitive acts, from the sorts of judgements that appear to be so obvious as to require no rationale (O-Intuition) to those that may be more subtle, or complex, but never-the-less present themselves to consciousness as being unsupported by an inferential process (H-Intuition). I have also argued that intuition corresponds to a kind of pattern recognition – that which is seized upon in the act of ‘judging that p’ is a pattern, and the judging itself is an act of bare recognition.

The definition of pattern that emerged turned out to require an operational flavour – patterns of sufficient complexity (highly mind- or context- dependent and not amenable to algorithmic interpretation) are patterns because they are *taken* as patterns by some pattern-recognizing creature. The creature “knows it when it sees it”, but independent identification conditions seemed pretty hard to come by. Vacuity in the definition was avoided by asserting that the relevant pattern-recognizing creature be embedded in the normative fabric of a socio-linguistic community.

The act of recognition itself seemed to require a computing primitive different in kind from the traditional Turing model. A series of abilities humans appear to exhibit in identifying patterns of interest – fringe consciousness, ambiguity tolerance and essential/inessential discrimination – were grouped together under the over-riding ability *perspicuous grouping*. I argued that the ability to perspicuously group events as relevantly similar was the basis of a new computing primitive *similarity-for-free*. 
In arguing that the Turing model is unable to provide for this ability - and connectionist networks demonstrably are so capable – I urged a commitment to taking the implementation story seriously and a commensurate acceptance of the repercussions that follow. Intuitive judgments were found to consist in a relevant subset of occurrent signals passing through the network. These intuitive occurrent signals are acts of bare recognition that are dependent on the set of weights and connections that comprise the network’s knowledge. Although every network is unique, the role of language and consciousness provides for a means by which individual networks are embedded in a socio-linguistic environment. That environment provides for normative constraints on the pattern-recognition behaviour of the network.

Taking the implementation story seriously provides a concrete way in which our concepts and our intuitions hang together\(^6\(^\text{44}\)\), and I have espoused a theory of concepts – pragmatic conceptualism – which parallels my analysis of intuition. I limit my analysis, however, to the original problem: what is the proper role of intuition within the larger context of SPM, in which concepts are deemed to be captured by analytic definitions, and intuitions are taken as evidence for or against those definitions? I shall argue for a proper role for intuition in the more general case, and shall treat epistemic intuitions as a sub-set\(^6\(^\text{45}\)\) of the general problem.

---

\(^6\(^\text{44}\)\) Note that much more is at stake than intuition itself when the implementation story is taken seriously. The structure of thought itself, the ways in which that structure is enabled, the role of language in that enablement, the relationship between the description of the high-level performance of a cognitive system and the causal methodology by which that performance is achieved: all of these basic cognitive stories are affected by taking the connectionist view seriously. I shall, of necessity, leave many of these philosophically fascinating stones unturned and focus on the particular relationship of intuition and conceptual analysis within the framework of SPM.

\(^6\(^\text{45}\)\) For those who regard ‘knowledge’ as a special case I respond by arguing that my treatment of intuition is general enough to be applicable to all concepts and all theories of concepts. I shall return to this question presently.
6.1 On The Proper Role of Intuition

The use of intuition is both ubiquitous and necessary within SPM (as I have defined it) and certainly ubiquitous in philosophical analysis in general - but it is also both fallible and highly variant. Intuitions may vary in any number of ways: over time within a single individual, among individuals and among cultures. Highly variant intuitions are antithetical to the enterprise of SPM as I have characterized it. Having explained that variance in Chapter 5, I turn now to the question as to how the variance and fallibility of intuition might be addressed.

Two items of interest will emerge from the analysis that follows. First, the role and degree of theory – both the implicit theory held in the form of expertise of the intuiters and the explicit theory against which intuitions must be balanced – is of central importance. Second, variance of network response may be characterized as a bias of the network, and must be taken into consideration. The proper role for intuition, then, is one in which it plays an essential - but highly moderated – role within a larger framework of conceptual investigation. The methodology governing that investigation must regard intuitions as empirical data that must be managed and balanced against other considerations - much like empirical data in any other discipline. Intuition, then, is not the magic bullet adherents of IDR or SPM wish it to be.

For some, however, intuition is to be ruled out-of-court even before the fallibility issue may be addressed. To those skeptics, I first turn.

6.1.1 The Intuition Skeptics

Hintikka is one of those who dismisses intuition altogether – it is a term that is meant to provide cover for what he sees as a hodge-podge of ordinary thinking and observation: “… what are explicitly called intuitions often turn out to be products of perfectly ordinary discursive thinking, combined with suitable observations.” It was Hintikka to whom I turned for an initial skeptical point of view at the start of this thesis: “For what is supposed to be the justification of such appeals to intuition? One searches the literature in vain for a serious attempt to provide such a justification.” Appeals to intuition are, for Hintikka, to be ruled out of court “unless the basis of such appeals is made explicit.” Williamson, too, provided a skeptical point of view with which I

---

646 Hintikka, 1999, pg. 143
647 Hintikka, 1999, pg. 130
648 Hintikka, 1999, pg. 147
opened Chapter 1: “… analytic philosophy has no agreed or even popular account of how intuition might work …”\textsuperscript{649}.

I can only ask Hintikka and Williamson to read Chapters 1 and 5 and judge whether such an account has been forthcoming. Leaving aside issues of ‘agreed upon’ or ‘popular’, it is just such an account I purport to provide and, in descending to the level of implementation in our brains, I could not be more explicit.

To be fair, although Williamson agrees with Hintikka that the term ‘intuition’ is really an unjustifiably concise term for what are a range of ordinary judgments, he is particularly concerned that the use of the term intuition is motivated by a desire to provide philosophy with a source of evidence that conveys the same credibility or claims the same evidentiary status as that of science. What Williamson wants to do is reject intuition as some special class of cognitive act that is so incontrovertible that it is to the content of the intuition – rather than the phenomenon itself – that we look. In wanting to talk about our \textit{intuitions} about some state of affairs, rather than the state of affairs itself, he contends that philosophers are attempting to get closer to their evidence, to refer to an evidentiary source that is incontrovertible and immediately apparent. Referring to intuitions about a concept, rather than the concept itself or even the referent of the concept, is meant to render philosophy empirically solvent – however, that move rends apart the phenomena under study and what is meant to be studied. The phenomenon under study becomes the intuition about a concept and what is meant to be studied is the concept itself. Referring to intuitions themselves opens a skeptical gap similar to that which exists between our sense perceptions and the world. Scientists have long ceased uttering claims like ‘I seem to see a magnet moving in a circular motion’ and instead claim something like ‘There is a magnet moving in a circular motion’, and Williamson argues that intuition must be similarly circumvented, or overshot, in order to close that gap. Williamson’s scepticism, then, is motivated by a view of intuition that both he and I reject: one that claims intuition is a special class of incontrovertible evidence such that it is to the content of intuition that we look to ground our investigations. Intuition is, for Williamson, better viewed as an ordinary judgment like any other, and as such gives us normal access to the phenomena we want to study.

To the problem of deep solipsism that results from the view Williamson rejects, I have no sufficient answer – nor do I think one may be forthcoming. Intuition is a cognitive act, and like any cognitive act, can always be accused of opening the door to some sort of horrible solipsism. Our interactions

\textsuperscript{649} Hintikka, 1999, pg. 130
with the world are mediated by our perceptual, nervous and neuronal systems - and it seems to me that the deep skeptic, the skeptic worried about solipsistic possibilities, will always remain unsatisfied as long as this mediation is admitted. That said, the connectionist account I provided in the previous chapter is an explicitly externalist account: concepts derive their semantic content from public discourse and worldly objects, as do intuitions. Our networks are trained by repeated exposure to objects and language. Code and process are explicitly linked on the C-Net account, and the world itself can be regarded as a part of that process, part of what shapes our plastic brains. Hence that by which our interactions with the world are mediated - our brains - can be said to be shaped by contact with that world. If the brain is shaped by the world, and it is through out brains that we access our concepts – can we not diminish these worries to a large extent? To the deep skeptic, one who is worried about the gap between experience and what is experienced, I can say no more.

However, to the more general skeptic – one who advocates the dismissal of intuition from the court of philosophical discourse – I put forth a challenge: given the explanation provided in Chapter 5, provide a means of delineating that which is to be rejected without also rejecting that which is the basis of much of the rest of our cognition. If intuition is the bathwater we want to throw away, I see no way to keep the baby.

If we are to dismiss intuition, without providing reasons to do so which clearly delineate that which we reject, then we are dismissing a category of cognition that shares a basic computational resource with everything from perception to category judgments to the understanding of natural language. Unless grounds for dismissal can be sufficiently delineated such that only those suspect intuitive judgments are included, then we are in danger of rejecting the very basis of our cognition. Even if I grant to the skeptic that intuition may be delineated both phenomenologically and that it may be delineated in the way in which I have done so in my analysis - that is not to grant that the reasons for dismissal have been delineated. No explanation for dismissal has been provided that does not entail a massive and injudicious pruning of our cognitive toolkit.

6.1.2 Intuition; Ubiquitous & Necessary, but Fallible & Variable

Recall Hintikka’s and Williamson’s dismissive assertions that intuition is used as a sort of ‘cover’ for other cognitive acts, as a misleadingly concise term for what are mere ordinary judgments. There is a way to interpret this claim that is less dismissive of intuition. While I have characterized intuition as a phenomenologically distinct sort of spontaneous judgment, I have provided a cognitive
explanation that is certainly meant to characterize it as *continuous* with all sorts of other cognitive acts. So while I agree that – from a cognitive mechanistic perspective – intuition may well be regarded as a quite general phenomenon, this is no reason to dismiss it. Indeed, it is a reason to regard intuition as a much more *ubiquitous* phenomenon than ordinarily thought, and this provides a reason to regard it as *necessary* (or, to differentiate it from the philosophically loaded term, *unavoidable*). While I may agree that its epistemic status is often in question – it is *fallible*, after all – it is not something we can so easily dismiss.

Pust has presented a view in which the ubiquity of intuition is brought to the fore. As we saw in Chapter 1, Pust sees intuition as a very general phenomena, permeating philosophy from the ‘particularist’ intuitions associated with Gettier cases and the like, the more ‘generalist’ intuitions associated with instances such as the self-evident nature of some moral theory, to the very broad ‘global’ intuitions associated with such broadly construed cognitive acts as wide reflective equilibrium. The use of intuition cannot be avoided, for Pust, as it permeates all the way up the hierarchy of philosophical cogitations, from the immediate and particular intuitions about Swampman and Gettier-cases, to the more general intuitions employed in reflective equilibrium.

Pust limits his purview to those intuitions strictly associated with philosophical discourse, and on this issue we part ways. I have extended that view and regard the everyday use of the term as continuous with the more philosophically austere definition. I have also posited that such diverse cognitive phenomena as perception, everyday reliable ‘hunches’, expert intuition, and even the conscious apperception of orderliness in one’s own train of thought, are all manifestations of the same broad underlying cognitive mechanism. In that sense, I am committed to an even more widespread use of intuition than Pust, but both he and I agree that the use of intuition is both unavoidable and ubiquitous.

If intuition is so widespread what, exactly, does it provide to us? For Goldman and Pust (G&P) it is infallible evidence of our own concepts – although perhaps contaminated with theory - and on this I agree. Hampered by a prior commitment to internalism, however, they are unable to provide a bridge to the external world, and are similarly unable to regard intuition as fallible. For me, intuition is infallible as evidence of our own concepts but fallible as evidence of concepts.

As we saw in Chapter 1, for Goldman & Pust (G&P) intuitions are to be regarded as reliable evidence – fulfilling the Reliability Indicatorship (RI) criterion - of the content of one’s own

[650] See Section 5.3.4.2
concepts, although that content can only be “a psychological structure or state that underpins a cognizer’s deployment of a natural-language predicate”. G&P are committed to a strictly internalist interpretation of mental content, and their account of intuition reflects this view. It can be seen, although my account is unabashedly externalist, that there is a clear line of agreement here. On my account intuition is an occurrent signal that is part and parcel of a much larger counter-factually defined set of signals that comprises a concept. Hence, intuitions are indeed to be regarded as evidence of one’s own concepts. Indeed, it is direct evidence as it is to be regarded as a context-dependent, temporally-fleeting instantiation of a concept.

We differ markedly, however, in that my concepts have a clear line of contact with non-psychological content – external content – by virtue of the role consciousness can play in associating concepts with publicly-accessible semantic content, concepts. The same can be said of intuitions and intuitions. Mine is an unabashedly externalist account. While intuition is infallible evidence of a concept, it is not an infallible evidence of concept.

It is the bridge to public semantic content that I have introduced that renders intuition fallible: just as one’s concept can be mistaken, so too can one’s intuition. I have likened the possession of a concept to that of a skill, the degree of possession of which is measured against successful interactions with one’s socio-linguistic community and the world at large. The counter-factually defined set of occurrent signals that comprises one’s own concept manifests itself in behaviour, including linguistic behaviour, which is an indication of how well that concept is in accord with concept.

Thinking a rock is edible is a mistake the world will quickly correct for me, and mistaking a sax for a trumpet is a mistake my peers will soon fix. It is this fallibility that must be addressed and accounted for – but this is no deep problem. People are fallible in any number of ways, but we manage to get on in the world. A similarly practical solution can be found for the use of intuition – and here, too, I part ways with G&P.

6.1.3 Ask the Experts; The Role of Theory

G&P are concerned with possible theoretical contamination of intuitive judgments. As noted earlier, G&P claim that “philosophers rightly prefer informants who can provide pre-theoretical intuitions

---

651 As quoted earlier on pg. 29.
652 See Section 5.3.4.1
about the targets of philosophical analysis, rather than informants who have a theoretical ‘stake’ or ‘axe to grind’\(^{654}\). Here the distinction between folk-intuitions and expert-intuitions is made clear. While it is clearly not an epistemic boon to solicit opinions, intuitions or other sorts of evidence from those with an ‘axe to grind’, that is a totally separate question from whether pre-theoretical intuitions should be sought or, to be perfectly blunt, whether that is even coherent. What is a pre-theoretic intuition? I’ve made it pretty clear that even perception is to be regarded as permeated through-and-through with conceptual content, and conceptual content begs to be placed within some – however simple –theoretical framework. An apple is seen to be an apple because I have learned that it tastes good, and fills my belly. I must hold a theory - however naïve - about food to see an apple as an apple (and that is how I see it). The very notion of pre-theoretic intuitions is incoherent.

From a computational perspective, the operation of any neural network automatically employs theory in the sense that – as we have seen – a network’s knowledge is to be regarded as residing in the set of weights and connections. Any activating signal makes use of that knowledge. Granted, the point at which that set of connections can legitimately be called theory is contentious, but only in the sense that the explanation of ‘theory’ has not been fully discharged. That criticism is valid, and is a weakness in any C-Net account. However, that there is a limited role for theory to play is clear, and given networks of sufficient complexity, inevitable. Theory pervades intuition. The quality and optimal extent of theory, in any particular case of intuitive judgment, is an open and interesting question – it is not a case of either/or.

It may well be true that in some cases we want to strip away as much theoretical baggage as possible in order not to pre-judge the case, or contaminate the judgment with prior (read: extraneous) theoretical commitments. So we may want to minimize theoretical contamination – that makes sense – but there still remains a minimum theoretical component. I take it that it is still the naïve folk from whom G&P want to solicit intuitions, since this is the only method they put forward by which one minimizes the invasion of theory. This makes sense if what is under study is what the folk think, but that is an enterprise distinct from that which strives to get at the content of some concept under study. Gopnik & Schwitzgebel (G&S) have pointed out that there is a category mistake in viewing the intuitions of the folk as evidence rather than data. Intuitions of the folk are data in the sense that they are to be explained – why is it that the folks think X and not Y? – but they are not necessarily evidence as to the content of some concept. The folks can be wrong and have been shown to be wrong (see Wisniewski, 1993).

\(^{653}\) See Section 5.4.3
\(^{654}\) Goldman & Pust, 2002, pg. 78
It is a tenet of IDR that it is in the minds of the folk – in the minds of everyone and anyone – in which we find the determinate content of our concepts, and it is through the use of intuition that we get at that content. But why, exactly, is it that the folk are the ones from whom we want to find out about matters epistemic? Shouldn’t we be asking a philosopher? Why would I ask the folk about concepts related to the space-time manifold? Shouldn’t I be asking a physicist? The two projects are distinct: the first is concerned with the sorts of concepts held by the uninitiated, the second is concerned with how those concepts might be refined with study. Theory ‘contamination’ is only an epistemic sin if it could be reasonably characterized as being equivalent to having an ‘axe to grind’ – otherwise it seems to me that holding a theory is a necessary component of almost all judgements. Recall that a concept held by a socio-linguistic community is only made precise – truthful, I want to say – by virtue of that community diligently investigating the issue at hand. Holding a theory is a part of that investigation. Experts within our community are what enable that community to develop much of the semantic content held by our language, and any individual possesses a theory relevant to a concept to a larger or lesser degree.

For proponents of IDR, and presumably adherents of Fodor’s Language of Thought, there is no distinction between novice and expert – any conceptual content which is to be found is to be found in the head (and in anyone’s head). IDR makes this claim explicitly, and I infer this conclusion from Fodor’s publicity requirement (see Section 4.4.3). I have argued that what is to be found in the head is a network, trained to a larger or lesser degree to pragmatically cope in the world, and respond appropriately to external and internal stimuli. The capability of a network to respond appropriately directly corresponds to the quality and quantity of training that network has undergone. In other words – expertise matters. This seems quite straightforward, and without some prior commitment to finding what you want in an untrained head, so uncontroversial as to seem obvious.

The fallibility problem is only partly addressed by probing the intuitions of those who know something of the matter at hand, however. Granted, one must be aware of theoretical bias – this I take to heart, and on that matter G&P have a point. But the matter is not so straight-forward as to eschew the expert an embrace the novice, nor is it a simple matter of ask the expert and ignore the novice - it is a balancing act, to be preformed carefully and with diligence, skirting naivety on the one hand, and dodging bias on the other.
6.2 Network Bias

I have argued that the role of theory may be played by the set of weights and connections in a network, and hence any intuition emerging as a pattern of activity in a network is permeated with theory. I want now to characterize the way in which any individual’s ‘theory’ of a domain might differ from some standard-bearer. Each network is distinct, reflecting the distinct background and set of causal interactions with the world, but each network—to the extent that they each share a culture and inhabit similar worlds—is similar. ‘Bias’ is meant to capture the idea that individual networks drift away from some norm. I am not proposing a general method of measuring the ‘uniqueness’ of a network; rather, I shall do so within the restricted domain of SPM. I am interested in how bias may be defined in this context and what factors cause or influence the degree of bias.

‘Bias’ implies a shift away from some norm or expected value. Instruments that exhibit bias are measured and calibrated relative to some standard. In this case, it is much more difficult to define bias in a manner that allows for calibration against some independent value. The normative ‘value’ against which I propose a network’s bias may be measured is what I have denoted concept\textsubscript{public}, which can be seen as encapsulating a ‘theory’ of the concept. However, under SPM the intuition itself is considered as evidence for or against some notional concept\textsubscript{public}, or can even be regarded as constitutive of the concept\textsubscript{public}. In other words, concept\textsubscript{public} is itself often tested, or generated, with use of solicited intuitions. Therefore, the notion of some completely independent standard of measure against which bias in the network is to be defined makes no sense. What shall emerge from this mutual dependence is a new sort of equilibrium between concept\textsubscript{public} and the intuitions\textsubscript{public} taken as evidence, established by a good methodological framework.

SPM assumes the prior establishment of a conceptual analysis—often a definition—which is being tested against imaginary (or real) cases. That pre-defined notion—which I here capture as concept(X)\textsubscript{public}—may be regarded as the social norm against which the network is to be tested. If corroborating output is identified by intuition(X)\textsubscript{public}, and manifested in the network as intuition(X)\textsubscript{mental}, then bias is meant to capture the cognitive forces within that network which can push or pull the occurrent signals into a different configuration, say intuition(Y)\textsubscript{mental}. For example, someone is asked whether a given case is an instance of concept(knowledge)\textsubscript{public}, and the corroborating response—the response that would confirm the definition under scrutiny—may be

\footnotesize{\textsuperscript{655} Section 5.4.2.2}
affirmative. In that case, bias may be regarded as cognitive forces that would increase the likelihood of generating intuition(not-knowledge)\textsubscript{mental} or intuition(just-believes)\textsubscript{mental}.

In this restricted sense, bias in a network is associated with ways in which the network may be regarded as ‘pushing’ or ‘pulling’ a signal away from performance in strict adherence with social norms or with the analysis put forth as the concept\textsubscript{public} under scrutiny. Bias may teach us something about the performance of the network – does the person really know what a ‘weak king side’ is? – or it may teach us something about the concept itself – the expert intuits that ‘knowledge’ has nothing to do with ‘justification’. The question of which lesson to learn – is the network bias an indication of poor performance or is it an indication that the social norm needs correction? – is a difficult one, and can only be addressed by an appropriate investigative methodology.

I shall analyze bias in a network from two perspectives. First, there are inherent biases in the network itself, which result from that network’s particular topology, hard-wired at some point in time. Second, there are (what I shall call) ‘externally provoked’ biases, which result from the manner in which the imaginary case is presented to the intuiter. These two forms of bias are related: externally provoked biases often rely on (or take advantage of) inherent biases.

My analysis of each of these biases assumes the notion of rampant context-dependence developed in Section 5.3.2. Transformed signals may be internally context-dependent, both temporally (sensitive to signals preceding that signal due to the recurrent nature of the network) and spatially (sensitive to concurrent signaling). Transformed signals may also be externally context-dependent, sensitive to the larger context in which an example is contained.

I shall speak here in very general terms, and use the Gettier cases under scrutiny to provide a more precise analysis in the next section.

6.2.1 Inherent Bias

A network sees what it knows, and can only see what it knows at any one time – this encapsulates what I mean by inherent bias. Networks must transform inputs into pre-existing categories, regardless of whether or not that category agrees with the concept\textsubscript{public} associated with that signal. Networks continuously evolve, certainly, but have a some set of pre-existing categories at any one point in time. One way of characterizing this bias follows the expert/novice distinction provided
above – novices will force categorization into known categories and do not have the discriminatory ability of an expert. However, in forcing transformed signals to fall into existing categories – under SPM a person is forced to make a choice, shades of grey or ambiguities are washed out by the way in which intuitions are solicited\(^{656}\) – there is an open-ended set of associations that may be brought to bear to perform the necessary disambiguation. Inherent bias is extended to include the set of pre-existing conceptual associations that may be used in order to force ambiguous signals into pre-defined conceptual spaces.

A network’s architecture dictates how it will respond to an external stimulus, and that architecture has a fixed topology at any one point in time. Recall that the topology of a network reflects the space of categories into which signals are transformed and the degree of discrimination among stimuli. New experiences are initially categorized into existing and known categories. Until one learns something about music, it’s all ‘happy’ or ‘sad’, ‘fast’ or ‘slow’ – one’s ability to discriminate patterns into more subtle categories is limited by the topology of the network. Ongoing training of a network can result in supervised or unsupervised learning, of course, and the architecture evolves, and new categories emerge. But at any point in time, the architecture and associated topology are fixed.

It is one of the trumpeted strengths of C-Nets that they are so tolerant of noise, deviations from standard prototypes, previously un-encountered situations, etc. and that strength requires that networks ‘see what they know’. That tolerance entails the ability to transform a wide variety of novel signals into known areas of stability – in other words, the tolerance for ‘error’ necessitates transformation of new signals into existing categories. If this were not the case, novel signals would constantly evoke novel categories and the stability required by any cognitive system would be absent. All signals would be tropes, and no signals tokens of a type.

In SPM a definition is thought to provide identity conditions on the conceptual category in question. When that definition is being tested against intuitive responses to some hypothetical case, those responses have a predetermined range – normally an affirmation or denial that the case is an instance of the concept. One is normally not permitted to say ‘sort of, but not really’. To avoid ambiguity, a novice network will presumably make use of any and all background knowledge at its disposal, regardless of whether or not that background knowledge is something the intuiter is aware of, or whether it would even be regarded as relevant by an expert.

\(^{656}\) This is not to imply that SPM cannot allow that some concepts are vague – only that for the person from whom intuitions are being solicited, the case is posed in a way that their response cannot be vague.
In the computational environment of a C-Net, such knowledge comes in the form of prior conceptual associations. These associations can be made active in the form of occurrent signals triggered by the external stimulus, and so be made part of the internal context brought to bear in disambiguation. In forcing me to respond in an unambiguous way, some portion of my background knowledge- which is uniquely mine - is brought to bear as a kind of internal bias.

Internal bias then, consists in both the pulling and pushing of signals into a pre-defined set of categorization functions, as well as the set of associations that provide for a disambiguating context.

6.2.2. Externally-Provoked Bias

Network response is sensitive to the external context, the larger context in which a stimulating signal is present, both the immediate context and the larger temporal context. Two dog encounters may be each processed into a transformed signal that is ‘dog’, but depending on the immediate context (barking, sitting, running, etc.) one encounter may also generate an associated signal of ‘scary’ and the other an associated signal of ‘clever’. A temporal context dependence may be regarded as an instance of the well-known ‘priming’ effect, in which a psychological bias may be explicitly generated by prior presentation of data (regardless of whether or not that data is something of which the observer is consciously aware, in which case the priming effect is a subliminal).

The point here is that one cannot easily separate the conceptual items of interest from the larger context in which they are presented to the network. An externally-provoked bias is one that takes advantage – either with intent or not – of the influence the larger context has on subsequent internal processing. In the constructed cases that WN&S generate, it is pretty clear that externally-provoked biases are being consciously manipulated in order to demonstrate the variance of intuitive responses. Detailed examples to this end are provided in the next section.

Externally-provoked biases are not independent of inherent biases, but are dependent on them. When a running brown dog is presented to the network, it may generate signals corresponding with ‘fear’ because of past encounters with brown bears. The signal itself does not contain the concept ‘fear’ as part of its context, but does so only indirectly in conjunction with the network that makes the brown-fear association, which itself generates the appropriate signal.
Understanding that a network’s response to a stimulus is context-dependent in these ways requires that intuitions be solicited in a methodical, rigorous way that attempts to either minimize or understand these biases. It is overly simplistic to think that one can simply construct examples or stories in ever more unrealistic and imaginative ways, to which one is to respond intuitively, and think that the semantic item being sought is itself unaffected by the imaginative example. The semantic item under scrutiny – knowledge, whether or not Swampman is conscious, the three-dimensionality of space, whatever – is tightly bound up with any number of other concepts that are rendered occurrent by the networks response to the example. The intuitive response being solicited can be – and often is quite *purposely* – shaped by the larger context in which the example is placed. There is a reason why Gettier-type examples become ever more complex – they are designed to solicit particular responses from particular groups. Rather than acknowledging the necessity of a rigorous, methodical elimination of bias, these examples are designed to take *advantage* of those biases. To these examples I now turn.
6.3 Variance of Epistemic Intuitions: WN&S

I return now to the original problem as posed by WN&S: if intuitions regarding matters epistemic are found to vary among cultural groups, then in what way may those intuitions be regarded as generating claims about knowledge with any real normative force? The idea that intuition solicited from the folk can be a tool that generates claims of normative force they dub Intuition-Driven Romanticism (IDR). I have positioned IDR as a sub-category of the much broader Standard Philosophical Methodology (SPM).

WN&S have constructed imaginative cases – which they call ‘intuition probes’ - in order to reveal cultural bias in epistemic intuitions. Their findings are not theoretical, but empirical; the intuition probes were given to real subjects, and the intuitive responses recorded. In that sense, I view their findings as empirical evidence in support of my general argument. My analysis of intuition can thus serve to explain the results they have found, buttressing their argument and generating more specific responses to their critics.

There are two separate hypotheses WN&S tested, each positing ways in which epistemic intuitions might vary: from culture to culture and from one socio-economic status (SES) to another. Partial results are given below. Two other untested hypotheses are that epistemic intuitions may vary as a “function of how many philosophy courses a person has had” and on the order in which cases are presented. Predictions as to expected results of these last two hypotheses are quite straightforward, given the preceding analysis. Indeed, in commenting on the possible relevance of the temporal order of presentation, WN&S allude to the path dependence of connectionist theory; “This hunch is reinforced by some intriguing work on neural networks suggesting that a variety of learning strategies may be ‘path dependent’. If this hunch is correct, the pattern of intuitions that people offer on a series of cases might well differ systematically as a function of the order in which the cases are presented.”

6.3.1 The Variance as Network Bias

Although WN&S provide a number of these intuition probes, the main point can be brought out by focusing on two examples corresponding to each of the two hypotheses actually tested. I analyze

---

657 WN&S, 2001, pg. 13
658 WN&S, 2001, pg. 13
each of these in terms of network bias. The first is a Gettier-style example, and was used to test variance in intuitions among Westerners and East Asians. The example is pretty standard Gettier stuff:

Bob has a friend, Jill, who has driven a Buick for many years. Bob therefore thinks that Jill drives an American car. He is not aware, however, that her Buick has recently been stolen, and he is also not aware that Jill has replaced it with a Pontiac, which is a different kind of American car. Does Bob really know that Jill drives an American car, or does he only believe it?

Western subjects answered in a typical fashion (approx. 75% answered “only believes”, and approx. 25% answered “really knows”). The results for East Asians are quite different (approx. 45% answered “only believes” and approx. 55% answered “really knows”). In accounting for this difference WN&S point to research done by Norenzayan and Nisbett that indicates “[East Asians] are more inclined than [Westerners] to make categorical judgments on the basis of similarity. [Westerners], on the other hand, are more disposed to focus on causation in describing the world and classifying things.” Since in many Gettier cases, including the one above, the evidence which causes the person to form a belief is false, East Asians are less likely to withhold claims of knowledge.

It is beyond the scope of my analysis to criticize the specific reasons as to why East Asians respond differently from Westerners. I am not going to make any claims whatsoever about the precise ways in which cultures may have different cognitive dispositions – all I want to claim is that there will be variance in intuitive responses among groups of people, and that variance will reflect the unique causal-historical set of interactions that identify those people as distinct cultural groups. Likewise, WN&S are not so interested in validating the explanatory claims about why there are differences between these two cultural groups, rather they are interested in using those hypothesized differences to guide them in constructing examples designed to take advantage of any differences; “we tried to construct some intuition probes that would tap into this difference.”

---

659 “The East Asian subjects were Chinese, Japanese and Korean. Some of the experiments were conducted in Asia, others used East Asian students studying in the United States or first and second generation East Asian immigrants to the United States. The Western subjects were Americans of European ancestry”, WN&S, 2001, pg. 10, footnote 19.
660 WN&S, 2001, pg. 19
661 WN&S, 2001, pg. 14
What is important to note is that the variations in epistemic intuitions “point to divergent epistemic concerns”\(^{662}\) that is, they are “both systematic and explainable”\(^{663}\). It is entirely to the point that the variation in intuitive response does seem to reflect a more general cognitive distinction. On my analysis, this corresponds to the way in which cultural norms are instantiated in broad architectural differences among networks. If the set of weights and connections is thought to encode background knowledge, then variation in cognitive performance is to be expected given distinct cultural background, and intuitive responses are a manifestation of that distinction.

An externally-provoked external bias is being consciously constructed on the basis of hypothesized internal biases. Those internal biases take the form of a set of associations that have been re-enforced by cultural assimilation. In the case of Westerners, presumably causation and classification have been associated, and it is that association that forces the C-Net into a classificatory judgment that withholds attribution of knowledge. Nothing hinges, for me, on whether the internal bias has been correctly identified. What matters is that WN&S have succeeded in constructing intuition probes that demonstrate network bias.

A second example called the ‘cancer conspiracy case’ shows massive variance between people of low socio-economic status (SES) and high SES\(^{664}\). The case is as follows:

It’s clear that smoking cigarettes increases the likelihood of getting cancer. However, there is now a great deal of evidence that just using nicotine by itself without smoking (for instance, by taking a nicotine pill) does not increase the likelihood of getting cancer. Jim knows about this evidence and as a result, he believes that using nicotine does not increase the likelihood of getting cancer. It is possible that the tobacco company dishonestly made up and publicized this evidence that using nicotine does not increase the likelihood of cancer, and that the evidence is really false and misleading. Now, the tobacco companies did not actually make up this evidence, but Jim is not aware of this fact. Does Jim really know that using nicotine doesn’t increase the likelihood of getting cancer, or does he only believe it?

Although this case revealed some difference among cultural groups (Westerners answered 90% “only believes” while East Asians answered 70% “only believes”) it showed significant

\(^{662}\) WN&S, 2001, pg. 32
\(^{663}\) WN&S, 2001, pg. 32
\(^{664}\) Years of education is used to make this distinction: Low SES is composed of those who reported that they never attended university and high SES if they reported one or more years of university.
difference between low-SES and high-SES groups. People of low SES were evenly split (50% and 50%) on the answer, while those of high SES answered over 80% “only believes”. Again, WN&S hypothesize as to the reasons why the two intuitive response profiles are so different (high SES subjects are more critical and accept “weaker knowledge-defeaters”\textsuperscript{665}, for example), but the truth of those conjectures is not as important as the data itself. The externally-provoked bias here comes, perhaps, in the form of a complicated set of counter-factual possibilities (if the company lied, would Jim be aware of it?) and an implied ability to judge the evidentiary status of public claims. The amount of university education one has been subjected to clearly affects one’s ability to sort through these issues, revealing the bias. This should hardly be surprising given the novice/expert distinction I have made so much of in my analysis.

The externally-provoked bias reveals the inadequacy of the novice network to make the necessary categorical distinctions. The even split on the low-SES group indicates that if there were ever a case of an intuition\textsubscript{mental} signal being forced into a category that doesn’t really fit, this is it. The correct answer seems to be neither yes nor no, but a resounding “maybe”. That answer is not allowed, and so the network is forced to settle in one state or another, using whatever resources it may. Since the response is evenly split, it’s hard to argue that the associations brought to bear are shared among the group, or even relevant to settle the issue at hand. Given some university education, however, and there may be more relevant associative relationships at play, or the network may have a better-defined topological space corresponding to ‘knowledge’, or both. Since the externally-provoked bias is toward complexity, to put it simply, is it any wonder that a more experienced network is better capable of making a distinction in this case?

The results as a whole in the WN&S paper hardly seem surprising to me, I’m a little unsure why the variance it reveals isn’t an expected variance. Our response to a situation – including that of knowledge attribution – is highly contingent on our experience. This is not a deep philosophical claim, but merely one that follows from everyday experience. While there are arguments as to why ‘knowledge’ may be of a special kind, and epistemic intuitions correspondingly privileged (see next section), I see no compelling reason to see why our concept of knowledge is not a recognized pattern of intentional behaviour like any other, and subject to all the usual vagaries and idiosyncrasies of human interpretation, issues of evidence and justification, and a whole host of other inter-related concepts. While knowledge is certainly distinct from mere belief – this is what truth and justification conditions add to the mix – it is still an intentional state about which observers will make judgments (isn’t that what motivates Gettier-type cases in the first place?).

\textsuperscript{665} WN&S, 2001, pg. 27
Constructing cases to take advantage of those contingencies, building externally-provoked biases, should be easy.

There are a number of possible objections to the WN&S data, but none detract from the central idea that I am pushing: intuitions will vary among groups, reflecting differences in the ways in which their networks have been trained.

6.3.2 Epistemic Variation – An Analysis

There are a number of objections to the ways in which WN&S both generate and interpret their data. None detract from my main argument that SPM cannot rely on unmediated intuitions of the folk for conceptual analysis.

One objection is to argue that these are simply the wrong intuitions to solicit. The proper intuitions to solicit are those that “are accompanied by a clear sense of necessity”\(^{666}\). WN&S respond by arguing that those intuitions which do “come with a clear sense of necessity and modal import – \textit{strong intuitions}\(^{667}\) are far from universal - so much the worse for soliciting intuitions from the folk. The objection really confirms that the folk are not the ones to ask!

On my analysis, I have argued for a different notion of necessity; rather than the modal notion introduced by WN&S I have argued that the degree of \textit{epistemic} necessity associated with an intuition is directly related to the degree of introspectively available support\(^{668}\). It is epistemic, rather than modal, necessity that provides the intuiter with a sense that it ‘could not be otherwise’. To the extent that there exists a Fully Supportive Set of reasons supporting the decision, then to that extent the intuition appears as epistemically contingent. Hence, the only intuitions that would qualify as \textit{strong intuitions} are those for which one or more of the conceptual associations required to force the network into a ‘yes/no’ response are unavailable to introspection. My argument to this end relied on the idea that without an introspectively available rationale for the judgment then there is not sense in which an inferential step may be regarded as potentially wrong - hence one simply could not see how it could be otherwise and the intuition would appear as epistemically necessary. There is no reason to suppose that the intuitions solicited from the low-SES group in the Cancer Conspiracy example do not appear to the intuiter as strong intuitions – they would be an instance

\(^{666}\) WN&S, 2001, pg. 33
\(^{667}\) WN&S, 2001, pg. 33
\(^{668}\) See Section 1.5.4 Introspective Support and Epistemic Necessity
of Real-H-Intuitions. This is empirically testable – simply ask the intuiters to rate the degree of necessity of their response.

Another objection is to argue that the task, as construed by the intuiter, differs according to their cultural background. The task is ‘filled-in’ in a culturally-specific way (Sosa 2008). Hence, it is not ‘knowledge’ that is being tested but some culturally-specific construal of the task. This general objection can take a number of forms; perhaps linguistic differences mean that specific words are interpreted in different ways, or variations in reasoning skills mean the task is interpreted differently. I take it that this objection is equivalent to my arguing that there are externally-provoked biases in play and to that extent I agree with the objection. Externally-provoked bias is indeed in play, and it is largely responsible for the variation in intuitive response profiles.

But to suppose that those biases may be eliminated is to presuppose that the concept of ‘knowledge’ can be separated and rendered independent from any and all relevant conceptual variation that exists between cultures and people. Isn’t that what soliciting intuitions is meant to either prove or disprove in the first place? Perhaps such cases can be constructed, but until that is demonstrated the variance of intuition remains a problem for the adherent of IDR. This is not to say that something cannot be learned from applying externally-provoked biases, and using the variance as one way to shed light on what might be some universally-shared conceptual core.

This raises another objection. While there may be variation in the intuitive response profiles, there are strict limits on that variation, which indicate that there may be some conceptual core that is universally shared. This may well be the case; indeed WN&S admit they “do not mean to suggest that epistemic intuitions are completely malleable or that there are not constraints on the sorts of epistemic intuitions that might be found in different groups. … there may well be a universal core to ‘folk epistemology’”669. This does not change the fact that intuitive responses vary –what it does imply is that such data can be used within a larger conceptual analysis methodology to try and reveal that conceptual core. Intuitions have a role to play in that methodology, but they are certainly not to be taken - unmediated by some further methodological considerations–as reflective in and of themselves of that conceptual core.

A final objection is that the proper intuitions to solicit are not those unconsidered intuitions of the folk, but “those that emerge after an extended period of discussion and reflection”670. I accept the

669 WN&S, 2001, pg. 31
670 WN&S, 2001, pg. 35
force and spirit of this objection. The novice/expert distinction that I have advocated and the notion that intuitive judgments play a role within a larger methodology both indicate that such reflective analysis is warranted. However, if intuitions are to converge the process can be denoted as $\text{intuition(knowledge)}_{\text{mental}} \converges \text{on concept(knowledge)}_{\text{public}}$. The intuitions will converge – if they converge at all - upon the concept held by the reflecting community at large. Only in agreement with the concept(knowledge)$_{\text{public}}$ is network bias eliminated, for there is no other standard by which to measure that bias, or even define the idiosyncrasies of individual networks as bias.

The question now remains: to which community’s concept(knowledge)$_{\text{public}}$ will epistemic intuitions converge? A key component of pragmatic conceptualism is that such convergence can happen, but it remains an open and interesting question as to the extent to which different cultures will converge upon the same concept(knowledge)$_{\text{public}}$. I remain agnostic on this question, although someone who holds to a LOT theory of mind, or adheres to the IDR view, or holds a theory of concepts in which concepts can be provided with necessary and sufficient conditions, cannot be agnostic on this point. I note, however, that convergence of conceptual content is derivative of the sorts of ways in which communities commit themselves to evidence, argument, and belief-formation and, unless these are to be made universal, there is very little chance that such convergence will take place. WN&S seem to advocate a similar view; “… in the absence of any evidence we don’t think there is any reason to suppose that the sorts of marked cultural differences in sensitivity to epistemic vectors that our experiments have demonstrated would simply disappear after reflection and discussion.”

To put this point in as provocative a manner as possible I shall contribute my own empirically-testable epistemic intuition probe. It is based on the ways in which different groups within our own culture interpret each other’s behaviour, form beliefs and accept evidence:

Tom is walking down the street, and encounters the well-respected local priest, who has been preaching to his large church for years. The priest sees Tom, seizes him, and exclaims excitedly “Tom, you’re to be married next year! What wonderful news!” Tom asks the priest how he knows this, and the priest claims “I had a vision last night. The Good Lord spoke to me of you, and in the morning I meditated with my Bible in hand. The scriptures confirmed my vision.” Tom does indeed get married that next year. Did the priest really know Tom would be married, or does he only believe it?

---

671 WN&S, 2001, pg. 35
I predict very different intuitive response profiles among groups of people, co-related to the number of church-seats per capita in the region in which the person spent their formative years. If we are to eliminate the epistemic intuitions of those who answer “really knows”, then the problem of intuition selection is pushed back to the sort of position Goldman advocates, namely, one pre-selects means of forming beliefs that are virtuous (empirical evidence, rules of logic, etc.) and eliminates those viewed as a vice (acceptance on faith, sacred texts, etc.). We revert back to the idea of selecting ‘experts’ in the field of epistemology and seeking consensus among those elite.

None of these objections deny the basic argument that I have been making - indeed, they all point to the use of intuitions as a sort of raw data that needs to be calibrated, corrected, balanced against other concerns, be they theoretical or empirical. Such is the proper role for intuition that I advocate.
6.4 Intuition and SPM

When we are asked to respond intuitively to an imaginative example, such as a Gettier-type case, our minds do not deploy a syntactic text with discrete semantic values in order to perform some sort of Turing-like computation, nor do they deploy some pre-existing definition by which to make the relevant comparison. When we perform the sort of cognitive task associated with these spontaneous judgments, our minds are seeking relevant similarities and they are doing so by leveraging the very distinctive computational resources of a neural network. Identifying many patterns of interest is often a matter of “knowing it when you see it”.

These intuitive judgments rest upon the same computation in kind as perceptual judgments, the sort of categorization functions of which the connectionist literature abounds, and which are well understood from a network-analysis point of view. Just as we recognize some sorts of perceptual (or visual) patterns and categorize objects in ways that correspond to our visual systems and our interests, so too do we recognize other, more complicated patterns – conceptual patterns – and categorize them in ways that correspond to the ways in which our networks have been trained to pick out relevant similarities. The extension of the basic principles involved in the well-understood perceptual patterns to the more contentious conceptual patterns was motivated mainly by having to take the computational resources of neural networks seriously, and that move was in turn motivated mainly by a denial of the obvious Turing alternatives.

An intuitive judgment is the output of a network, stimulated by an input signal, the weights and connections of which represent the background knowledge that network brings to bear. Intuitions will vary among people just to the extent that their background knowledge varies. If intuitions are to be used as evidence in philosophical analysis – and I have argued the use of intuition is both ubiquitous and necessary\textsuperscript{672} - then they must be regarded as only one piece of evidence. I have provided reasons as to why intuitions may vary, what remains to be done is to develop a methodology that takes these variations into account. SPM as it is currently practiced is totally devoid of the requisite methodological considerations in taking at face value the intuitions of the folk.

\textsuperscript{672} I characterize the use of intuition as necessary in the case of SPM since it is a part of the very definition of that particular method of philosophical analysis. Whether philosophical analysis, more broadly construed, can take place without intuition seems unlikely, particularly given the very broad construal of intuition defined within this thesis – from a bare recognition of fact to the more generalist intuitions deployed in reflective equilibrium.
There is general agreement, however, among many philosophers, that the use of intuitions must be balanced against other considerations. Typical of these is Weatherson, who states that “the true theory of knowledge is the one that does best at (a) accounting for as many as possible of our intuitions about knowledge while (b) remaining systematic.” Presumably ‘remaining systematic’ relies on evidential sources other than intuitions themselves, and prior theoretical or methodological commitments are among them.

Accounting for the biases inherent in the use of intuition – both inherent and externally-provoked biases – requires a methodological rigor that views the solicited intuitions and the cases presented as part of a larger body of evidence to be weighed and balanced, as evidence so often must be. The profile of those intuited (novice or expert?), a careful examination of the cases being presented in order to bring out the background contextual fabric, the order in which cases are presented, existing theoretical considerations: these are all relevant to the careful methodology required for good conceptual analysis being done via the solicitation of intuition.

Bias may only be eliminated to the extent that the cases presented for intuitive response elicit responses in tune with what the socio-linguistic community has decided is the proper response – recall that the only standard to which bias may be measured is the concept public. Bias elimination seems a pointless exercise, unless one merely wants to construct cases designed to confirm what you already know. By provoking bias, though, it appears there may be much to learn – both about the concept being analyzed and the network being prodded to respond.

It seems to me that if SPM is committed to both analytic definitions and the solicitation of folk intuitions in response to ever more imaginative cases, then it is in a hopeless muddle. Given the deep contextual-dependence of network response, one could, I suspect, always construct some imaginary example that provoked inherent biases such that the proffered definition fails. The uncompromising nature of an analytic definition seems at odds with the individualistic and highly contingent nature of neural networks responding to cases of open-ended imaginative scope.

If SPM is to remain committed to analytic definition, then the solicitation of intuitions must be carefully managed and not always assumed to provide prima facie evidence for or against the definition. If SPM is to abandon the rigour of the analytic definition, and instead allow for a much more nuanced approach to the meaning of our concepts – including accepting the notion that some concepts correspond to an open-ended set of events the extension of which may only be specified

---

673 Weatherson, 2007, pg. 7
by the participation of an observer – then perhaps the analytic definition may be regarded as a robust initial hypothesis, to be balanced against the intuitive judgements of those who have gained a degree of familiarity with the concept in question. In any case, the intuitions of the folk and the analytic definitions of the philosopher make for unhappy bedfellows.
BIBLIOGRAPHY


Davidson, D., Essays on Actions and Events, 1980, Oxford University Press.


Goldman, A., Liaisons, 1992, MIT

Goldman, A. (& Pust, J. Chapter 4), Pathways to Knowledge, 2002, Oxford University Press


Hofstadter, D., I Am a Strange Loop, 2007, Basic Books

Horgan, T. & Tienson, J., Short Precis of Connectionism and the Philosophy of Psychology, 1999, in Acta Analytica 22

Kirsh, D., When is Information Explicitly Represented?, 1991, in Information, Thought and Content, ed. Hanson, P., UBC Press

Lakoff, G. and Johnson, Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought, 1999, Basic Books

Leibniz, Selections, ed. Philip Wiener, 1951, Scribner

Marr, D., Vision, 1882, Freeman Press


MacDonald, C., & MacDonald G., Connectionism: Debates on Psychological Explanation, 1995, Blackwell Publishers


Merleau-Ponty, Sense and Non-Sense, 1964, Northwestern University Press


Place, U., Connectionism and the Resurrection of Behaviourism, 1995, in Acta Analytica 12
Place, U., Connectionism and the Problem of Consciousness, 1999, in Acta Analytica 22

Polanyi, M., Experience and Perception of Pattern, in Modeling the Mind, ed. Sayre, K. & Crusson, F. 1968, Simon and Shuster

Pust, J., The Use of Intuitions as Evidence in Philosophy, 2000, Oxford University Press


Quine, W., Word And Object, 1960, Cambridge & MIT Press


Rand, T., Connectionism and Folk Psychology, 2000, Masters Thesis submitted to University of London


Sellars, W., Empiricism & The Philosophy of Mind, 1997, Harvard University Press


Weatherston, B., What Good are Counterexamples?, 2003, in Philosophical Studies 115, pg. 1-31


