In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Philosophy

The University of British Columbia
Vancouver, Canada

Date May 02/2003
Abstract

Fred Dretske (1981) claimed that the essence of the kind of cognitive activity that gives rise to Intentional mental states is a process by which the analogue information coming from a source-object is transformed into digital form. It is this analogue-to-digital conversion of data that enables us to form concepts of things. But this achievement comes with a cost, since the conversion must involve a loss of information. The price we pay for the lost information is a proportional diminishment in our ability to discriminate the source-object from others that may be similar to it. I argue that this fact underlies an important distinction between what a mental state may be about and to what the state may be directed.

Aboutness and directedness are two of four Intentional dimensions on which this project concentrates. The other two are aspectual shape and misrepresentation. The distinction between aboutness and directedness is a part of a proposed approach to Intentionality based on the script theory of Roger Schank and Robert Abelson (1977). Scripts are schemata—organized knowledge structures that guide our understanding of the world around us. Schank and Abelson's basic ideas are extended to yield four different script-types: episodic (related to situations and events), instrumental (related to procedural knowledge), personal (representing an agent's goals and plans), and definitional (involved in object-recognition).

The relationship between scripts and the Intentionality of thought is the main focus of this dissertation. An important secondary concern is the viability of externalism and internalism. It is argued that neither of these attitudes is independently adequate to provide a full account of Intentional content. Rather, the proper approach is to confine externalistic influences to aboutness and then characterize directedness in a manner that captures the world-according-to-the-agent. This strategy is implemented in the following way: aboutness is construed causally-evolutionarily; directedness is constructed with the help of the notion of an equivalence class; aspectual shape is shown to be a function of the kind of information a script provides; and an account of misrepresentation is given by comparing the different extensions generated from aboutness and directedness respectively.
# Table of Contents

Abstract .................................................................................................................. ii  
Table of Contents .................................................................................................... iii  
List of Tables .......................................................................................................... vii  
List of Figures ......................................................................................................... viii  
Acknowledgements ................................................................................................. ix  
Dedication ................................................................................................................ x  

I  Introduction  
Project Purpose and Scope .................................................................................... 1  
1.1 Intentionality and its Significance ...................................................................... 1  
1.2 Intentional Instrumentalism ................................................................................ 3  
  i. Carnap .................................................................................................................. 3  
  ii. Quine ............................................................................................................... 4  
  iii. Dennett .......................................................................................................... 5  
  iv. Comments ...................................................................................................... 7  
  • Inference to the Best Explanation ....................................................................... 9  
1.3 Externalism and Internalism ............................................................................... 10  
1.4 Objectives ........................................................................................................ 12  
1.5 Overview ......................................................................................................... 15  

II  Intentionality  
An Extended Definition ........................................................................................... 17  
2.1 Two Distinctions ............................................................................................... 17  
  i. Intentional States and Mental States .................................................................. 18  
  ii. Intentional States and Consciousness ................................................................. 19  
2.2 Four Intentional Dimensions ............................................................................ 20  
  i. Aboutness ....................................................................................................... 20  
  ii. Aspectual Shape ............................................................................................ 23  
  iii. Directedness .................................................................................................. 25  
  iv. Misrepresentation ......................................................................................... 29  

III  Skin, Boundaries and Authority  
Locating Intentional Content .................................................................................. 31  
3.1 Background to Externalism ............................................................................... 31  
3.2 The Role of the Environment ......................................................................... 37
i. Physical Furniture: Adam and Adam
ii. Sociolinguistic Conventions: “Arthritis” and “Tharthritis”
iii. Causal History: Swampman

3.3 Consequences of Externalism
i. Type-Identity Must be Abandoned
ii. Token-Identity Must be Abandoned
iii. Weak Supervenience Fails
iv. Strong Supervenience Fails
• Global Supervenience
v. Mind-Brain Reductionist Programs are Out of Reach
vi. Mental Causation is Jeopardized
vii. No Privileged Access to First-Person Mental States

3.4 Three Individualist Offerings
i. Fodor
ii. Searle
iii. Segal
• Natural Kinds and Social Externalism
• Two-Factor Accounts
• Radical Internalism

3.5 Descriptions, Motleys, and Neologisms

IV Scripts I: Foundation and Structure
And a Word About the Game

4.1 The Game

4.2 Conceptual Dependency
i. The Research Methods of AI and Psychology
ii. The Axioms of CD
iii. Primitive Acts
iv. Numeric Values and Scales
v. Representing Causation

4.3 Scripts: Episodic, Instrumental, Personal
i. Long-Term Memory (LTM)
ii. Episodic Scripts
iii. Instrumental Scripts
iv. Personal Scripts

4.4 Definitional Scripts
i. Inanimate Objects
ii. Animate Objects
• Animal Scripts
• Human Scripts
• Is There a “Me” Script?

4.5 The Basic Structure of Scripts
i. Headers and Primes
ii. Tracks
### V Scripts II: Evidence from AI and Psychology

5.1 Computer Task Modelling
- SAM
  - The English Analyzer and the Script Applier
  - The Summarizer
  - The Question-Answer Mechanism
  - The English Generator
- FRUMP
- TALESPIN
  - ELIZA
- Modelling Cognition

5.2 Support from Psychology
- The Semantic-Episodic-Procedural Dissociation in LTM
- Schemata and Scripts

### VI Thoughts on Physical Implementation

And Another Word About the Game

6.1 The Informational Origins of Thought
6.2 Naturalizing Scripts
- Object Recognition and Simple Definitional Scripts
- On the Precarious Existence of Water-Stains
- Complex Definitional Scripts
- Scripts\text{\_E}, Scripts\text{\_P}, and Scripts\text{\_P}
  - A Hierarchy of Physical Events
6.3 Priming by Autoassociation
- The Hebb-Marr Model of the Hippocampus
- Priming by Description
6.4 Original Intentionality, Holism, and the Frame Problem
6.5 Summary

### VII A Script for Intentionality

7.1 Aboutness, Directedness, and Aspectual Shape
- Intentional Aboutness
- Intentional Directedness
  - Designatory and Indicatory Rigidity
  - Fictional Objects
  - Privileged Access
- Aspectual Shape
  - Twin-Earth Cases
List of Tables

Table 4.1  A list of the D-goals used by Schank and Abelson (1977)  .................  87
Table 4.2  A list of the planboxes associated with each of the five D-goals  ...............  88
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 2.1</td>
<td>A representation of the relationship between mental states, conscious states, and Intentional states</td>
<td>20</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>A summary diagram stressing some of the points covered in chapter II</td>
<td>30</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>A Conceptual Dependency representation of the explicit and implicit causal connections in a script-based story</td>
<td>77</td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>A Conceptual Dependency representation of the generic RESTAURANTes</td>
<td>81</td>
</tr>
<tr>
<td>Figure 4.3</td>
<td>An example of a cognitive association between two scripts</td>
<td>95</td>
</tr>
<tr>
<td>Figure 4.4</td>
<td>A schematic summary of script theory</td>
<td>103</td>
</tr>
<tr>
<td>Figure 5.1</td>
<td>Overview of SAM's Question-Answer mechanism</td>
<td>110</td>
</tr>
<tr>
<td>Figure 5.2</td>
<td>Anatomical regions of the brain</td>
<td>120</td>
</tr>
<tr>
<td>Figure 5.3</td>
<td>A substructure of “The Czar and His Daughters”</td>
<td>130</td>
</tr>
<tr>
<td>Figure 5.4</td>
<td>Examples of pictures in simple and complex forms as used by Pezdek and Chen (1982)</td>
<td>131</td>
</tr>
<tr>
<td>Figure 6.1</td>
<td>A representation of the ACQUISITION hierarchy</td>
<td>148</td>
</tr>
<tr>
<td>Figure 6.2</td>
<td>A schematic of an event memory as a pattern of cell activations in the neocortex</td>
<td>150</td>
</tr>
<tr>
<td>Figure 6.3</td>
<td>A representation of Marr's model of the hippocampus</td>
<td>152</td>
</tr>
</tbody>
</table>
I gratefully acknowledge the assistance of my supervisory committee—Dr. Steve Savitt (Chair), Dr. Paul Bartha, and Dr. Mohan Matthen. Their comments and suggestions have made the final draft of my dissertation much more refined than it would have been otherwise. I am also grateful to Dr. Philip Hanson, who has helped to shape the initial stages of this project.

My sincere appreciation must also go to Dr. Manal Guirguis-Younger, Dr. Alastair Younger, Mohab Guirguis, and Michael DeSousa, who all have, in their own way, contributed to my studies. Mohab and Manal have especially been generous with their time, and have lent support and encouragement whenever it was needed.

Derek Walker has read every bit of this work and has made many corrections. I am not unmindful of the difficulty of editing a several-hundred page document and am happy to acknowledge his efforts.

Finally, I owe much to Wahba Ramzy, my grandfather, who has been there form the very beginning, and who has made himself familiar with every university campus I have ever attended.
For My Parents

Maurice and Laila Guirguis
Introduction
Project Purpose and Scope

Fortunate are they who can perceive the causes of things ... but more fortunate still are they who can bloody well perceive what to do about them.

Philip J. Davis, Thomas Gray: Philosopher Cat, p. 82.

1.1 Intentionality and its Significance

Our lives contain many activities. Some of these activities are non-mental; others are mental. An example of a non-mental activity is opening a door; an example of a mental activity is thinking about opening a door. In Psychology from an Empirical Standpoint, Franz Brentano set out to distinguish mental and physical acts, and to devise a proper method of investigating the former. He first considered the following proposal:

It would be possible for us to characterize physical phenomena easily and exactly in contrast to mental phenomena by saying that they are those phenomena which appear extended and localized in space. Mental phenomena would then be definable with equal exactness as those phenomena which do not have extension or spatial location. (1874:84)

Brentano was not wholly satisfied with this analysis, however. His chosen distinction was that, unlike physical phenomena, “every mental phenomenon includes something as object within itself” (1874:88). He called this “something” the Intentional in-existence of an object.

As a preliminary formulation, then, we might say that Intentionality is that property of some mental states by which these states are represented as being directed toward, or about, or of objects, events, or states of affairs. The difference between Intentional aboutness and Intentional directedness is a crucial one, and will be discussed in some detail in the next chapter. For now, we just note that any mental state with an Intentional character takes something specifiable as its object. I interpret “Intentional object”—or simply, “object”—broadly to include living organisms, inanimate items, events, states of affair, etc. If, for example, I have a belief, it must be a belief that such-and-such is (was, will be) the case; if I have a fear, it must be a fear of something, someone, or some circumstance; if I have a desire, it must be a desire for someone, something, or that some event should come to pass. And so on through a large number of cases.\(^1\)

\(^{1}\) In Principles of Philosophy, Descartes goes as far as equating material objects with extension or space. “The nature of body,” he claims, “consists not in weight, hardness, colour, or the like, but simply in extension” (II:4); and a little later, “there is no real difference between space and corporeal substance” (II:11).

\(^{2}\) One thing should be noted immediately: “intending” and “intentions” in the sense of someone planning to take a certain action, or the road to hell being paved with good intentions, is just one among many types of mental states that are...
Chapter I  Introduction: Project Purpose and Scope

This initial definition suggests that at least some mental acts—such as those involved in believing, fearing, desiring, and deciding—depend on a subject matter or content. Hence one cannot decide unless one has some content to decide upon; one cannot believe or desire unless there is something to believe or desire. Brentano was not only aware of this fact, but took it as the basis for differentiating the mental and the physical realms. Yet the exact nature of the content in question—what it means for a psychological state to have a subject matter—has been elusive. In fact, it is not an exaggeration to say that contemporary philosophy of mind has been preoccupied with delineating what Intentional content amounts to, without achieving much consensus on the matter.

The quest for an adequate theory of Intentionality is not just a popular pursuit; it is a vitally important one. And indeed, much of the literature on the topic has an unmistakable tone of urgency to it. Thus Jerry Fodor tells us that producing a naturalistic theory of Intentionality is an essential step in vindicating common-sense psychology, for “if common sense Intentional psychology really were to collapse, that would be, beyond comparison, the greatest intellectual catastrophe in the history of our species” (1987a: xii). Fred Dretske, speaking more broadly of the special insight people seem to have into why they do the things they do, claims that to give up such authority would be to “relinquish a conception of ourselves as human agents. This is something that we ... will not soon give up” (1988a: x). These are harsh warnings, indeed. Why all the fuss?

There are many reasons why people might want a theory of Intentionality, but among these various motives one stands out: concern about irrealism with respect to the mental has been a central theme in the philosophy of mind, and producing an adequate theory of Intentionality is seen to be a crucial step in the debate. Fodor claims that for many people the most worrisome fact about Intentionality is its ontological oddness:

[T]he deepest motivation for Intentional irrealism derives not from ... relatively technical worries about individualism and holism ... but rather from a certain ontological intuition: that there is no place for Intentional categories in a physicalistic view of the world; that the Intentional can't be naturalized. (1987a: 97)

We are nothing more than physical creatures, and this simple fact raises a genuine problem for Brentano’s enterprise. His thesis—which Dennett (1978: xvii) calls the “workhorse” of the philosophy of mind—alleges that while all mental phenomena are Intentional, no physical phenomenon exhibits that property. This has been traditionally taken to be an irreducibility hypothesis: the mental, in virtue of its Intentionality, cannot be reduced to the physical.

But now consider: if mental states are just certain sorts of activities occurring in certain kinds of brains, then a mental act is no more than a physical act, and one of much the same type as that involved in opening a door, sneezing, or swallowing a bit of beef. Brentano’s “Intentional inexistence”—the basis of his mind-body division—turns out to be as much a physical property as mass, volume, and heat conductivity. Since only some physical states are “Intentionally blessed,” as one might say, the problem now shifts from separating mind and body to producing a theory of what makes certain physical (e.g., brain) states Intentional. In the above citation, Fodor is expressing what has come to be known as the ontological problem. More generally, he is expressing concerns regarding the success of post-Darwinian reductive approaches to the mind, and about what a failed reduction of Intentionality might entail for the integrity of folk psychology.

characteristically Intentional. “Intentionality” in its more common usage ought not be confused with the technical sense we are employing here. In order to distinguish the technical sense from the ordinary sense, I shall follow Searle's convention of capitalizing the technical occurrences throughout this work (including quoted passages).
What makes the ontological problem unsettling is that we do know one possible consequence of an unsuccessful reduction: folk psychology might be declared a false theory, a relic of an old, unsophisticated conception of the mind, retainable only for the sake of convenience. *Instrumentalist* arguments of this kind have already been proposed, most famously by Rudolf Carnap, W. V. O. Quine, and Daniel Dennett. To get a better idea of what mental realists are up against, a small detour to consider such views seems warranted.

### 1.2 Intentional Instrumentalism

According to William Lyons, "to suggest ... that the new reductionist approaches to Intentionality ... were simply the result of the behaviourist victories over Cartesianism in both philosophy and psychology would be an oversimplification ... At least as important as the rise of modern behaviourism was the retooling of nineteenth-century positivism by the Logical Positivists of the Vienna Circle" (1995: 9). *Positivism* is the doctrine that the only genuine empirical knowledge is scientific knowledge, and the only genuine means of gaining knowledge is the scientific method of producing and testing causal hypothesis by means of observation and experimentation. Positivism was thus a "hymn to natural science," extolling it as the sole repository of facts, while, at the same time, repudiating metaphysics and all its works.

The *logical positivists* were the obvious offspring of this movement. An important part of what was new to this logical version of positivism was its belief that logic was the chief tool of philosophy, just as mathematics was the chief tool of science; and the idea that the task of philosophy was to work on the logical and conceptual foundations of science. This task was held to comprise the reduction of the language of the human sciences to the language of physics. With respect to psychology, the sought reduction entails the translation of folk psychological language into physical language: initially into neurophysiological or behavioural language, but with the expectation that ultimately it would be possible to translate neurophysiological statements and statements about behaviour into statements couched in the language of physics. One of the central figures involved in this program was Rudolf Carnap.

#### i. Carnap

In "Psychology in Physical Language," one of Carnap's background doctrines is that any true science must ultimately be expressible in physical language, for only a language made up of terms for observable objects, properties, and events will be truly universal and intersubjective, and so truly scientific. Because (strictly speaking) there are only physical events, only statements about physical events could be literally true or false. In the context of psychology, this means that in the absence of a mature neurophysiology, we must fall back on behaviouristic physical idioms. As Carnap (1933: 165) put it, "all sentences of psychology describe physical occurrences, namely, the physical behaviour of humans and other animals. This is a sub-thesis of the general thesis of *physicalism* to the effect that *physical language is a universal language*, that is, a language into which every sentence may be translated."

---

3 Interestingly, Carnap's outlook did not start out that way. He initially sought to reduce all knowledge to what he called *phenomenalistic language*: "I believed that the task of philosophy consists in reducing all knowledge to a basis of certainty. Since the most certain knowledge is that of the immediately given, whereas knowledge of material things is derivative and less certain, it seemed that the philosopher must employ a language which uses sense-data as a basis. In the Vienna discussions my attitude changed gradually toward a preference for the physicalistic language" (1963: 50).
Carnap does not demand that psychology reformulate each of its sentences in physical terminology. On the contrary, he encourages psychologists to use whatever vocabulary they like, just as long as the production of definitions through which psychological language is linked to physical language is possible. In this case, the generalized sentences of psychology would be translatable into physical vocabulary, and the laws of psychology would be physical laws. So, “now it is proposed that psychology, which has hitherto been robed in majesty as the theory of spiritual events, be downgraded to the status of part of physics” (1933: 168).

Carnap was setting an agenda for philosophical psychology in which only questions of language would be featured. Viewed through an even more widely angled lens, Carnap was suggesting that traditional philosophy rid itself of all metaphysics and concentrate fully on the logical analysis of the language of science (1932: 60-1).

ii. Quine

While Carnap did not leave an account of Intentionality as such, his views on how philosophical psychology ought to proceed had significant impact on modern discussions of the subject. Among the many who were influenced by Carnap was W. V. O. Quine. Even though Quine has never embraced the logical positivism of the Vienna Circle—and in his “Two Dogmas of Empiricism” has expressly rejected the analytic-synthetic distinction embedded in the logical positivists’ verification principle—Quine’s view of psychology is surprisingly close to Carnap’s.

Like Carnap, Quine believes that physics is the fundamental science, and only what is sanctioned by physics is part of the true and ultimate structure of reality. Quine argues that his ontology—what he takes to be bedrock reality in the universe—precludes mental or Intentional items such as beliefs and desires. Of course, we may find it useful in our ordinary lives, even indispensable in practice, to talk in terms of beliefs, desires, intentions, and the rest, but we should not be misled by the utility of Intentional idioms into thinking that they describe what is really there. As Quine himself put it in Word and Object (which, incidentally, was dedicated to Rudolf Carnap, “teacher and friend”):

One may accept the Brentano thesis either as showing the indispensability of Intentional idioms and the importance of an autonomous science of Intention, or as showing the baselessness of Intentional idioms and the emptiness of a science of Intention. My attitude, unlike Brentano’s, is the second ... Not that I would forswear daily use of Intentional idioms, or maintain that they are practically dispensable. (1960: 221)

Quine sometimes puts his views about psychology in terms of a difference between intensional (with-an-s) and extensional language. The extension of a term is whatever real object, property, or relation—or in general what fact of the matter, if any—is usually (that is, conventionally) picked out, or referred to, or individuated, or selected by the use of a sign or symbol

4 "This neutral attitude toward the various philosophical forms of language based on the principle that everyone is free to use the language most suited to his purpose, has remained the same throughout my life. It was formulated as 'principle of tolerance' in logical Syntax and I still hold it today ...” (Carnap 1963: 18).

5 A simple version of the verification principle claims that a statement (or sentence) is literally meaningful or significant if and only if it is either empirically verifiable (or falsifiable) or it can be seen or shown to be true (or self-contradictory) simply by means of the analysis of the conventional meanings of the signs or symbols used in the statement. The meaning of all terms and phrases is thus anchored to checkable facts about either language or the world. Something like this version of the verification principle is found in Carnap (1963: 45) and Ayer (1946: 35).
of a language. The intension of a term is its meaning, or sense, or significance for any user of the term. Thus for Quine the correct language or notation for fundamental natural science is extensional, because an extensional language homes in directly on what is real without an intervening subjective understanding or viewpoint.

**iii. Dennett**

Perhaps the clearest version of Intentional instrumentalism in the Carnapian-cum-Quinean mode is that of Daniel Dennett, whose first book, *Content and Consciousness*, contains the following revealing passage:

The content one ascribes to an event, state or structure is not, then, an extra feature that one discovers in it, a feature which, along with its other, extensionally characterized features, allows one to make predictions. Rather, the relation between Intentional descriptions of events, states or structures (as signals that carry messages or memory traces with certain contents) and extensional descriptions of them is one of further interpretation. If we relegate vitalist and interactionist hypotheses to the limbo of last, desperate resort, and proceed on the assumption that human and animal behavioural control systems are only very complicated denizens of the physical universe, it follows that the events within them, characterized extensionally in the terms of physics or physiology, should be susceptible to explanation and prediction without any recourse to content, meaning or Intentionality. (1969:78)

One can certainly hear the echo of Carnap and Quine in these remarks; for Dennett is telling us that what is really there—bedrock reality, as it were—can be captured fully by an extensional vocabulary. Any deployment of Intentional idioms arises in answer to some felt need for “further interpretation,” or a heuristic overlay of some kind upon the facts.

Dennett develops his account of Intentionality in *Brainstorms*, where he seemed to embrace a kind of functionalism. Ned Block (1980a) identifies three senses of “functionalism”: analytical functionalism, where a system is explained in terms of the capacities of its parts and the way the parts are connected to one another; computation-representation functionalism, a special case of analytical functionalism that emphasizes the “computer-as-mind” analogy; and metaphysical functionalism, the hypothesis that mental states simply are functional states. Seen against the background of Block’s distinctions, Dennett’s functionalism seems somewhat idiosyncratic. It is driven only by pragmatic considerations: in a way that produces useful predictions about how humans will behave in given circumstances, Dennett believes that we can attribute Intentional functional states to them. But in so doing, he warns us to resist any inclination to think that we are thereby picking out or individuating real, detectible brain states or processes. It is a mistake to expect that attributing similar beliefs to two or more people would lead to our finding similar brain states or processes operative in the persons concerned; for we attribute functional states to individuals, not on the basis of neurophysiological knowledge, but on the basis of how they behave in light of what they can be supposed to have perceived in the environment.

Dennett argues that our Intentional descriptions of human behaviour can be seen as a particular attitude or stance, the Intentional stance, which people take up to their fellow humans, to other animals, and at times even to machines. We deliberately view our neighbours (our pets, our cars) as functioning in terms of a belief-desire system; and we do so because taking up the Intentional stance is a useful and economical way of predicting what humans will do.

But there are other stances beside the Intentional. For instance, we might consider a machine from the point of view of its design—that is, we might adopt a design stance. In taking up this
attitude one is considering an object from the point of view of what it has been designed to do (or not
do), not from the perspective of what it might think or want. Alternatively, we might contemplate a
thing from a physical point of view. To take up the physical stance is to consider something insofar
as it is made up of certain materials with certain properties. Accordingly, to adopt the physical stance
in regard to humans is to investigate their physiology, their chemistry, or—at a more basic level—
their physics. For Dennett, a migration from common-sense Intentional explanations and predictions
to the more reliable design-stance explanations and predictions (and eventually to the explanations
and predictions of the physical stance) is the “proper direction for theorists to take whenever
possible” (1978: 12).

Dennett describes his next book, *The Intentional Stance*, as “a series of post-Brainstorms
essays in which I attempted to revise, re-express, and extend my view” (1987fc: ix). The result is a
work with a markedly increased emphasis on the pragmatic nature of Intentionality, and on the
assertion that its value lies almost wholly in its power to predict behaviour. “The perverse claim
remains: *all there is* to being a true believer is being a system whose behaviour is reliably predicted
via the Intentional strategy” (1978: 29).

On this more uncompromising view, a thing can be said to have Intentional states only in the
sense that, when we attribute information-carrying or content-containing states to it, these
attributions enable us to predict what behaviour will be produced by such a system. Beyond this, the
mind has no Intentionality for the simple reason that there is no mind, and the brain has no
Intentionality because it has neither states nor processes with content. Intentionality is merely a
feature of our language: it just so happened that our ancestors (through a process of trial and error)
evolved an Intentional way of talking about one another that enabled them to predict behaviour with
success.

Dennett seems to have followed Quine in holding that either we naturalize Intentionality by
a downward reduction of folk psychology—first to behaviourism, then to neurophysiology, and
finally to physics—or we take the view that it will be impossible to succeed in this downward
reduction and acknowledge the fiction of Intentional language. If reduction is possible, then it would
have to be a total reduction—including those phenomenal or subjective features of our mental
lives—in the context of a materialism akin to what Galen Strawson refers to in the following
statement:

> Serious materialists must hold experiences to be physical phenomena in every respect, and hence even in
> respect of their having the experiential character they have ... It follows that they cannot talk of the physical as
> opposed to the mental or experiential at all ... The distinction that concerns them when it comes to the mind-body
> problem cannot be a distinction between the mental and the physical, because it is a distinction that must be
drawn entirely within the realm of the physical. If one is a materialist, to say that there is a fundamental distinction
> between mental and experiential phenomena, on the one hand, and physical phenomena, on the other hand, is
> like saying that there is a fundamental distinction between cows and animals: that on the one hand there are
cows, and on the other hand there are animals. (1994: 71)

The difference between total reduction and full-blown elimination is that, in the former case,
it is expected that some future “scientific” psychology would reduce the common-sense conception
of mentality if a sufficient number of statements licensed by the scientific psychology meshed with,
or matched up with, or correlated with statements licensed by the common-sense view; otherwise the
common-sense conception would have to go. The kind of reduction involved would be an instance of

---

what Ernest Nagel (1979) called heterogeneous reduction. If we let $T_R$ (reduced theory) designate the theory or set of experimental laws to be reduced, and $T_B$ (basic theory) designate the theory to which the reduction is effected or proposed, $T_R$ would, in the present context, be folk psychology and $T_B$ would be physics.

Nagel argues that in cases of heterogeneous reduction, the distinctive traits that are the subject matter of $T_R$ sometimes demonstrably fall into the province of a theory that has been initially developed to handle qualitatively different materials, and so may not include some of the descriptive terms or concepts of the to-be-reduced theory in its own set of theoretical statements. The basic theory might thus appear to eliminate distinctions familiar within the reduced theory, thereby producing a sense of mystification, especially if $T_R$ deals primarily with macroscopic phenomena, while $T_B$ deals primarily with microscopic phenomena.

As far as the reduction of folk psychology is concerned, however, Nagel observes that no sense of mystification is warranted, since a successful heterogeneous reduction would only remove the need for folk psychological idioms; it would not eliminate folk psychological categories. He explicitly makes this point using headaches as an example:

... the reduction of one science to a second ... does not wipe out or transform into something insubstantial or "merely apparent" the distinctions and types of behavior which the secondary discipline [i.e., $T_B$] recognizes. Thus, if and when the detailed physical, chemical, and physiological conditions for the occurrence of headaches are ascertained, headaches will not thereby be shown to be illusory. On the contrary, if in consequence of such discoveries a portion of psychology will be reduced to another science or to a combination of other sciences, all that will have happened is that an explanation will have been found for the occurrence of headaches. But the explanation that will thus become available will be essentially of the same sort as those obtainable in other areas of positive science. (1979:366)

Viewed in this way, the (heterogeneous) reductionist position is that the linguistic representational classifications of folk psychology will have no place in a mature theory of human action. So if we use the label "cognitive science" as a catchall for the various scientific disciplines that will play a role in the explanation of human behaviour, then what the reductionist is claiming is that the Intentional vocabulary of folk psychology will be expunged from a fully developed cognitive science. The stronger eliminativist position is that our common-sense mental states do not exist: there are no such things, just as there are no such things as phlogiston, caloric fluid, and witches on broomsticks. Essentially, eliminativism denies that our folk psychological categories have the reality required for any sort of reduction to take place.

Reductionism and eliminativist do not therefore entail one another. It is perfectly consistent to simultaneously be a reductionist and a realist with respect to the mental. Carnap, in concerning himself only with reducing the sentences describing human behaviour to the language of physics, may be viewed as a reductionist in the heterogeneous sense. On the other hand, one can be an eliminativist by rejecting the reality of the mind without thereby endorsing reductionism—without, that is, rejecting the practical utility of folk psychological talk; this we see in Quine-cum-Dennett instrumentalism.

iv. Comments

A few remarks are in order. First, I think the threat to folk psychology from theories that deny the reality of Intentionality accounts for much of the urgency on the part of mental realists to find a workable account of Intentional content. Even those who promote irrealist viewpoints
(Dennett included) understand why their arguments are unpalatable. Folk psychology is well
engrained in all of us, and is not likely to vacate the philosophical premises without a quarrel.
Moreover, some believe that folk psychological explanations are probably the best we can hope for.
Thus Stephen Pinker remarks,

"I call an old friend on the other coast and we agree to meet in Chicago at the entrance of a bar in a certain hotel
on a particular day two months hence at 7:45 P.M. I predict, he predicts, and everyone who know us predicts that
on that day at that time we will meet up. And we do meet up. That is amazing! In what other domain could lay
people—or scientists, for that matter—predict, months in advance, the trajectories of two objects thousands of
miles apart to an accuracy of inches and minutes? And do it from information that can be conveyed in a few
seconds of conversation? (1997:63-4)"

"It is this predictive power which is most problematic for instrumentalism. Dennett maintains
that all there is to being a true believer is being a system whose behaviour is reliably predictable via
the Intentional strategy. But that our Intentional talk gets predictions right, though it is just a useful
story with no hold on the real facts of the matter, seems just too improbable to be true. It seems too
fortuitous for credibility that the folk psychological account of the sources of human action is not
merely neat and comparatively easy to understand, but also possessed of enormous predictive power,
yet at the same time cannot be said to be a true, factual picture of anything.

It is not helpful to claim that "the fact that we are products of a long and demanding
evolutionary process guarantees that using the Intentional strategy on us is a safe bet" (Dennett
1987b: 33). Though an appeal to evolution might explain why we use the Intentional stance—
presumably because it has served us well in the past—it does not explain why the Intentional stance
works; it does not shed any light on why there should exist any predictive strategy that is at once
powerful, successful, and unrelated to the facts at issue. Something is missing, for evolution could
not produce successful predictions based on any stance unless it had also brought it about that the
stance in question had some causal connection to actual states of affairs—some kind of causal
relation, proximate or distant, to the behaviour one is trying to predict. This is why no matter how
many adjustments we make, no variation on, say, the astrological stance—which lacks the requisite
connection to human behaviour—will ever produce a predictive strategy that has anything more than
chance accuracy. The success of the Intentional stance thus suggests that there must be more to the
matter than a blend of pragmatism and blind luck.

There is also a question of whether Dennett’s instrumentalism can avoid the implication that
human heads have real contents. When we talk about some person or organism via the Intentional
stance, we attribute to the person or organism a belief that so-and-so, or a desire that such-and-such.
It is the “so-and-so” and “such-and-such” that are at the heart of our Intentional talk and ultimately
what gives it predictive power. So, for instance, if I see a man in front of a restaurant, pacing back
and forth, looking intermittently at his watch then down the street, I can say, with reasonable
confidence, that the man is waiting for someone to join him, that he believes that the person who is
to join him is not in the restaurant and is late, that he is annoyed by his friend’s tardiness. What is
interesting is that I have interpreted the man’s behaviour in terms of what information he himself
seems to have gathered by means of his senses, and the behavioural response that he has made to that
information.

Generalizing this observation, we can say that the interpretations we make about someone,
and the subsequent predictions we form regarding how that person will behave, depend on our

7 See also Guirguis (1999).
knowledge of his or her information—that is, on what we can learn about the perceptual information in that person’s “input slot” and his or her forward planning as revealed by relevant signals and reactions. Putting our folk psychological descriptions in this way makes it clear that they imply acceptance of the claim that the human brain is an information-processing or content-utilizing system. Folk psychological predictions are successful because ordinary folk got it right: between input and output the human brain operates as a device that processes information.

I do not offer these criticisms as decisive, but I think they help to clarify our options: we can either accept instrumentalism and live with the apparent conclusion that the predictive power of folk psychology is merely an evolutionary piece of good fortune—perhaps brought about by a parallel evolutionary process in theory construction—or we can reject instrumentalism as it stands. The former option, though possible, seems to me unlikely: the chance of any progressive process of constructing theories hitting upon a successful method of predicting human behaviour—a method that bears no relation to actual facts—seems remote. So, committed to the claim that there is more to Intentional content than pragmatic utility and coincidence, I choose to reject instrumentalism as it stands.

**Inference to the Best Explanation**

Some of the comments I have made may be taken as invoking a version of the so-called inference-to-the-best-explanation argument, which has often been urged in the debate between scientific realism and constructive empiricism. This debate was initiated by Bas van Fraassen, whose critique of scientific realism reached a wide audience with the publication of his book, *The Scientific Image*. Roughly, a scientific realist holds that (i) science aims to give us, in its theories, a literally true story of what the world is like, and that (ii) the acceptance of a scientific theory involves the belief that it is at least approximately true. Van Fraassen espouses a version of antirealism that he calls constructive empiricism. Essentially, this view substitutes “empirical adequacy” for “truth” in the realist definition. It holds that (a) science aims to give us theories that are empirically adequate, and (b) acceptance of a scientific theory involves the belief that the theory is empirically adequate. A scientific theory is “empirically adequate” if it gets things right about the observable phenomena in the world, where the content of the predicate “observable” is to be fixed relative to our sensory abilities. Realists have countered by arguing that scientific realism provides the best explanation of the power to predict phenomena with which science provides us. This power would be a highly improbable cosmic coincidence if the theories were not true, or if the entities posited by those theories did not actually exist. A related position is convergent realism, the thesis that although any particular theory may be partially true, successive theories provide better approximations to the truth. Convergence realists see the history of science as one of progress toward real facts.

Some antirealists—e.g., van Fraassen—criticize inference-to-the-best-explanation by pointing to the underdetermination of theories by evidence: different incompatible theories may enable the same predictions. Where one is successful, so will be its empirical equivalents. He also argues that in the predictive success of science generally, there is no miracle to explain: our best theories are successful because we systematically reject those that are not. In a similar vein, Larry Laudan (1981) has attacked the convergence argument by stressing that the truth of a theory is not a good explanation of its predictive success, for there have been many past theories that were successful in their time—e.g., the phlogiston theory of combustion and the caloric theory of heat—that we now know to be false and referentially empty. Even when intertheoretic retention takes place, it only occurs with regard to a few selected elements of the older theory. Laudan concludes from his inspection of the evidential record that, far from supporting realism, the history of science supports
the opposite induction: the falsity of previously successful theories provides good reasons to think that presently accepted theories are also likely to be false.

This is but a brief sample of the debate. I do not have anything significant to add, but I do want to make the following brief points.

Suppose \( Q = \{p_1, p_2, p_3, \ldots, p_n\} \) is a set of phenomena for which we need an account, and \( T \) is a theory that provides our best explanation of \( Q \). I doubt that in practice either scientists or laity infer the truth of \( T \) only on the basis of \( T \)'s constituting our best explanation of \( Q \), since we can imagine situations where \( T \) may fall short of accounting for all the phenomena in \( Q \); for example, \( T \) may account only for \( R \), where \( R = \{p_1, p_2, p_3, \ldots, p_m\}, m < n \). In such cases, any reasonably rational individual would consider \( T \) to be at best incomplete, and would not be satisfied until \( T \) was somehow altered or augmented so as to achieve a "better fit" with the data. We do, however, tend to infer the truth of a theory when it is (to put it in van Fraassen's terms) empirically adequate—that is, when the theory accounts for all the phenomena or data that need explanation (i.e., when \( T \) accounts for all \( p_i \in Q \); that, I take it, is what "empirically adequate" means). But then the truth of a theory is inferred from its ability to explain the entire data set, and not just from its being the best explanation we have at a given time.

It is here that van Fraassen's point about the underdetermination of theories by evidence becomes important, for there may be a number of theories, incompatible with \( T \) and each other, that are nevertheless empirically adequate with respect to \( Q \). There is no question that this sort of underdetermination is relevant to the realist-antirealist debate, but how much does it impact folk psychology? Not very much, I think. Right now, folk psychology is a theory without a rival: we currently have nothing on offer which comes close to folk psychology's accuracy, power of prediction, and simplicity. Nor, for the same reason, do I think that Laudan's pessimistic induction is especially threatening: not only does folk psychology presently have no rival, but we have good reason to suppose that the theory has been around since pre-recorded history without undergoing significant changes. It was there, in its full bloom, while other theories came and went, and it is still here today. In fact, folk psychology might just be the most successful theory we have ever had. So the question of intertheoretic retention (or lack thereof) has simply never been an issue.

None of this suggests that constructive empiricism is untenable, nor is it meant to. Even if constructive empiricism turns out to be true, the consequences for folk psychology will be negligible. We must keep in mind that constructive empiricism is a hypothesis that affects all theories which posit unobservable entities (in van Fraassen's sense of "unobservable"), and this covers just about all of the natural sciences. So if it finally turns out that folk psychology is merely "empirically adequate" rather than "true," so be it. It will be in good company. As to how things stand right now, we are still left with the original choices: we can accept instrumentalism along with the conclusion that folk psychology is a pleasant coincidence of evolution, or we can reject instrumentalism as it stands. I still choose the latter course.

1.3 Externalism and Internalism

In light of the above discussion, I shall proceed on the assumption that Intentionality is a real property of some brain states—a property that, in my opinion, still needs a naturalistic explanation.
that will do it justice. Accordingly, one of the questions which concern my project is the viability of two rival approaches to Intentional explanations: externalism and internalism. Psychological externalism is a thesis that has its roots in semantic externalism. On classical theories of meaning—e.g., those of Frege and Russell—the semantic properties of some words are at least partly determined by the internal states of a speaker, by concepts or descriptions in the speaker’s head: for Frege, the sense of a word was so determined (see below, §3.1); for Russell, it was the meaning of universal terms and the constituents of descriptive propositions.

Russell argued that in addition to our acquaintance with particular things, we are acquainted with universals or concepts, and that every complete sentence must contain at least one universal. He believed that “we have acquaintance with anything of which we are directly aware, without the intermediary of any process of inference or any knowledge of truths” (1912: 25). This was the basis for what he characterized as the “fundamental principle” in the analysis of propositions containing descriptions: “every proposition which we can understand must be composed wholly of constituents with which we are acquainted” (1912: 32). Keeping in mind what Russell meant by “acquaintance,” his principle comes down to the claim that, every proposition understood by an agent is composed wholly of constituents whose meanings are directly understood by the agent, without the intermediary of any process of inference or any knowledge of truths. In short, Russell held that we must attach meaning to the words we use if we are to speak significantly and not utter mere noise, and that the meaning we thus attach is something with which we are directly aware.

This familiar picture, which was the target of Kripke’s famous Naming and Necessity lectures, eventually found a strong opponent in another picture of meaning as determined by external contextual conditions. On this new model, the reference of names and natural-kind terms (e.g., “marmoset”) is determined by external facts and states of affair. In using a name $n$, for example, a speaker $A$ refers to $x$ in virtue of the fact that $n$ traces back, through one speaker after another, to the individual $x$. It was quickly alleged that these externalist considerations could not be confined to language alone, and eventually, semantic externalism gave rise to psychological externalism: the doctrine that the Intentional character of mental states also cannot be independent of environmental determinants.\footnote{The evolution of externalism is briefly reviewed in Stalnaker (1993).}

Drawing a spatial boundary between two categories of physical events and states—those located inside an individual’s skin and those taking place outside the skin\footnote{This is not the same sort of boundary on which Ryle’s “official doctrine”—or, as he somewhat more abusively calls it, the “dogma of the ghost in the machine” (1949: 15-16)—is based: it isn’t a boundary separating the “physical” and the “mental,” where the former is supposed to be located in space and time, but the latter is supposed to exist in time and not space. Rather, the bifurcation I have in mind is completely within the physical domain (for there exists nothing else), and the boundary in question is the outermost limits of our bodies.}—we can say that psychological externalism is the thesis that the Intentional content of an agent’s mental states is not independent of the conditions in the agent’s environment; that Intentionality partly depends on physical states, or circumstances, or processes, or principles, or conventions located outside the agent’s body. In contrast, internalism—or individualism, as it is sometimes called—is the view that the Intentionality of states of mind depends on the intrinsic physical state of the agents possessing them.\footnote{A much more detailed analysis of the internalist-externalist debate is given in chapter III, where it is explained just how Intentionality is thought to be dependent on environmental conditions by the externalists and independent of such conditions by the internalists.}
In my judgment, neither of these two views is *independently* satisfactory, for neither takes account of facts that strongly suggest the plausibility of the other. What is required is a treatment of Intentionality that is sensitive to *all* Intentional dimensions. This, in a nutshell, is the task I have set for myself.

1.4 Objectives

My objectives are not independent of my motivations and some strong intuitions I have about Intentionality. I am dubious about the kind of reasoning that regards agent-environment relations as *the* determinants of Intentional content. External relations, while constituting an important part of cognition, do not tell the whole story. What is missing is the point of view of the agent—what is “in the agent’s head,” if you prefer. I suggest that *narrow content* is at least as important in determining the Intentionality of a given mental state as any external relations that state might bear. But, as we shall see, one gets different results depending on whether one chooses to individuate Intentional states by agent-environment relations (externally, broadly, or widely) or by reference to the agent’s perspective (internally, individualistically, or narrowly).

I intend to show that Intentionality is multidimensional—that it has both a wide or external dimension and a narrow or internal dimension—and that, rather than being an expression of one or the other, it encompasses them both. That is not to say that both components are equally important. They aren’t: Intentionality does not require that there always be a relation between an agent and an external object, but it does seem to require that the agent have a point of view or perspective *in virtue of* the mental state (s)he is in.

I call the broad dimension of Intentionality *Intentional aboutness* and the narrow dimension *Intentional directedness*. I take it that the most plausible naturalistic explanation of the former is an evolutionary one, such as has already been proposed by Dretske and Millikan. But evolutionary explanations have their limits, the most important of which is their inability to account for the directedness aspect of Intentionality. Consequently, giving a naturalistic account of Intentional directedness is an important goal of my work. I propose to accomplish this goal by combining Dretske’s information-processing approach with some basic resources from *script theory* as developed by Roger Schank and Robert Abelson in *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures*.

I first encountered scripts in a cognitive psychology seminar and immediately realized that they had the potential to coherently combine all my intuitions: that mental states supervene on physical states, that physical doppelgängers must (therefore) be psychologically indistinguishable, that Swampman is not a zombie, that the perspective of the agent contributes essentially to his or her psychology. The key to seeing all this is to separate the agent’s point of view from his or her relationship to the environment—to acknowledge, in other words, that Intentional states have a *directedness* as well as an *aboutness*.

Even though the theory of scripts has been somewhat underutilized by philosophers, I believe that it can be developed to answer my basic goals. Ultimately, then, I wish to promote a

\[^{12}^\text{Narrow Content, which is supposed to capture the agent’s perspective or point of view, is here contrasted with wide content; see chapter II for details.}\]

\[^{13}^\text{We will discuss Millikan’s account in chapter III, Dretske’s in chapter VI.}\]
script theory of Intentional content, which, as far as I know, has some claim to originality. The ideas I borrow from script theory will be first delineated and then reworked into an information-processing mould. Here are some highlights of the position I intend to defend:

i. I uphold the sense-reference distinction for Intentionality, which is to say that Intentionality has both an aboutness (reference) and a directedness (sense). Whereas an Intentional state may not be about anything, no Intentional state is undirected.

ii. Aboutness and directedness are mutually independent Intentional dimensions, in that neither determines the other. The former is fixed by a causal relation, \( C \), in which an Intentional state stands with an object. This relation may be mediated just as long as the aetiology traces back to an actual item. Directedness, on the other hand, is fixed by the indicatory-identificatory function of scripts.

iii. Scripts can be grounded in a Dretskean information-processing theory. In particular, scripts emerge as a result of the transformation of information from analogue to digital form.

iv. Whether or not an Intentional state is about an exclusive object depends on what lies at the other end of the \( C \) relation. But whether or not a mental state is directed toward an exclusive object depends on what the agent, as an information-processing system, is capable of discriminating under a specific script. Aboutness and directedness can therefore give rise to different extensions. What goes into the aboutness extension (\( E_a \)) of an Intentional state is what stands in \( C \) to the relevant representation; what goes into its directedness extension (\( E_d \)) is determined by the primed script and the discriminatory resources of the information-processing system whose script it is. By comparing the aboutness extension and the directedness extension of an Intentional state, we can account for many common types of error or misrepresentation.

v. The question of whether Intentionality is to be defined internally (as a property of brain states and their relations) or externally (as a relation between agents and world) can be answered as “neither” or “both,” depending on one’s point of view. Intentionality is not homogeneous; it has multiple facets that demand different analyses. Thus the aboutness component of Intentionality is externalistic and relational and ought to be treated as such, whereas the directedness component is likely to resist such treatments. If this is correct, any attempt to account for Intentionality exclusively in one way or the other—in terms of either individualism or externalism—can attain partial success at best.

vi. I want, nevertheless, to leave the door open for relativizing the manner in which we individuate Intentional states to specific theoretical aims and objectives. Without denying the Intentional dimensions outlined above, I see no reasons why one cannot focus on a single Intentional feature (and ignoring others), if by so doing a certain theoretical question or avenue of investigation is highlighted. Under specific conditions, then, the decision of how to individuate Intentional states may be a pragmatic one. I talk a bit more about this in §2.2.iii.

vii. One must be wary of the word “script,” since it can be misleading. Scripts are not linguistic models, at least not necessarily linguistic. For the purposes of this project, I shall adopt a relatively broad interpretation of “script,” one that is consistent with the mainstream psychological conception of a schema. This means that insofar as schemata can be non-propositional, scripts can be as well. In fact, scripts represent many different knowledge
Chapter I  Introduction: Project Purpose and Scope

domains. There are scripts for personal stereotypes and roles, scripts for goal oriented actions and common event sequences. There are scripts for spatial scenarios, personal habits, objects, animals, and persons. Moreover, scripts may—and, as we shall see, often do—include visual and acoustic information, olfactory, gustatory, and other “purely phenomenal” data. In all cases, the content of a script will be highly structured, not simply a list of features or properties.

viii. In chapter IV, several distinct types of scripts will be introduced, most of which will not be based on verbal information (though they may be primed by words and sentences). And in chapter V, the fact that many scripts are non-linguistic will be supported by brain-imaging studies revealing the relative autonomy and non-semantic nature of the long-term memory systems housing various kinds of script-based knowledge. Several terms have been given to name these structures: “schemata,” “stereotypes,” “themes,” “macrostructures,” “models,” “frames,” and “memory organized packages.” Any of these designations would do for our purposes, but I use “scripts” to preserve continuity with Schank and Abelson’s (1977) research in which that particular name was used.

ix. Many scripts, but not all (especially those representing people; see §4.4.ii), are generic in that they provide general information about the components, attributes, and relationships that typically occur in their exemplars. This makes the generic script for, e.g., eating at a restaurant different from the specific memory trace constructed when an individual eats at a particular restaurant at a particular time. The relationship between generic scripts and specific memory traces is laid out by the script-copy-plus-tag (SC+T) hypothesis in §5.2.ii.

x. It is convenient to view scripts as having “slots”—or frames, as I shall say—which are filled as the script guides the processing of a specific input (Minsky 1975). For example, the frames of a restaurant script include character frames (customer, cook, waitress, hostess); object frames (tables, chairs, food, menus); and action, plan, or goal frames (the customer orders food, the cook prepares the order, the waitress serves the food, the customer eats). These frames are filled with contextually specific information when someone comprehends a particular restaurant experience.

xi. A distinction will be made between two stages of script utilization: script identification and script application. During the identification process the comprehender “searches” for the script that provides the best fit for the input. This is essentially a process of pattern recognition. As information accrues in a data-driven fashion, the information matches the components, attributes, and relationships of one script better than others. Once a script has been “identified” or invoked, the application stage starts, during which the script guides processing in a conceptually-driven manner. Several phenomena occur during script application. First, the script influences the perception and interpretation of the input material; experiences would be ambiguous or difficult to understand if no script provided background knowledge. Second, the script governs the attention that is allocated to the elements in the stimulus. In most conditions more attention is devoted to information that deviates from the script than information that is relevant to the script (this is also explained by the SC+T hypothesis). Third, scripts play an important role in inference generation, a process that (normally) occurs when frames are filled by default. Finally, scripts provide the knowledge base for formulating expectations about subsequent events during comprehension.
Chapter I  Introduction: Project Purpose and Scope

How scripts licence inferences, despite the fact that they are (or could be) non-propositional, is a matter that depends on what we mean by “inference.” There is no question that the process of comprehension benefits greatly from non-verbal information, and that numerous assumptions—based entirely on extra-linguistic facts—are being constantly made: a man comes in from outside with a dripping umbrella, and immediately we assume that it has been raining. What is the nature of that assumption? If it is an inference, then clearly it is one that is made without any linguistic cues. On the other hand, if we insist that an inference is strictly a relation between sentences, then (trivially) many scripts cannot be said to generate inferences, in which case we shall have to give another name to the kind of assumption made from the wet umbrella (the reader is free to use his or her imagination here). But I prefer a more inclusive interpretation. We normally assume in a restaurant setting that a cook prepared the food we ordered, even though we never see this task. I consider this kind of assumption to be an inference, and will call it so, without thereby committing myself to the claim that the script which licences it is a linguistic construct.

1.5 Overview

There are many approaches to Intentionality. Other than the instrumentalist approach discussed above, we find: the linguistic approach derived from the work of Noam Chomsky (1968) and exhibited most fully in the work of Jerry Fodor (1975, 1987a); the teleological approach developed by Ruth Garrett Millikan (1984, 1993) and Colin McGinn (1989); the information-processing approach embodied in the work of Fred Dretske (1981, 1988a); the functional role approach of Brian Loar (1981); and the more recent developmental approach recommended by William Lyons (1995). Although these different perspectives are all important, only some of them are directly connected to my position. I shall, therefore, restrict the next two chapters to presenting only the background necessary to the exposition of my thesis.

In chapter II, I provide a somewhat extended definition of Intentionality. The distinction between Intentional states and mental states, on the one hand, and Intentional states and consciousness, on the other, will be discussed. The multidimensional character of Intentionality will also be emphasized—including such Intentional dimensions as aspectual shape, aboutness, directedness, and the possibility of misrepresentation.

The controversy concerning the notion of privileged or direct access to first-person mental states serves to focus some of the motivations behind individualism and externalism, along with various specific positions these two stances have inspired. This is the target of the third, largely expository chapter. Special attention will be given to the different ways in which the environment has been claimed to influence Intentional content, including physical furniture and background (Hilary Putnam), social and linguistic practices (Tyler Burge), causal history (Donald Davidson), and evolutionary development (Ruth Garrett Millikan and Colin McGinn). These views will be contrasted with the internalist doctrines of John Searle, (early) Jerry Fodor, and Gabriel Segal. Included in my discussion are the implications of an externalist psychology to psychophysical supervenience, mind-body identity, and the causal efficacy of the mental.

As scripts are discussed almost exclusively in psychology and artificial intelligence circles, a basic introduction seems to be an appropriate theme for chapter IV. I start by outlining Conceptual Dependency, an ancestor of script theory. I then go on to differentiate various types of scripts—episodic, instrumental, personal, and definitional—and define the meaning and function of headers, frames, priming effects, default values, and other related notions.
Chapter V is where I survey some of the most important evidence for scripts gathered in the last two decades. The chapter will be divided into two sections. In the first, I take up the kind of support which involves task modelling in the field of artificial intelligence. The power and versatility of script-based computer programs like SAM, FRUMP, and TALESPIIN become obvious as these systems are contrasted with non-script-based programs like ELIZA. In the second part of the chapter, I turn to the psychological evidence, starting with neurological support for the semantic-episodic-procedural distinction in long-term memory, and ending with a description of some experiments that directly test the predictions of script theory in social settings.

I intend scripts to account for the directedness dimension of Intentionality. But in order to do so in a non-question-begging way—that is, in order for scripts to ground the Intentional property of directedness in a manner that does not itself presuppose Intentionality—they must be shown to have a non-Intentional foundation. Accordingly, chapter VI outlines the details of Dretske’s information-processing theory, and shows how Dretske’s notion of analogue-to-digital data conversion (which he uses to explain concept formation) provides the requisite foundation.

In chapter VII, the scriptal approach will be tested against the characteristics of Intentionality outlined in chapter II: aspectual shape, aboutness, directedness, and error or misrepresentation. I deal with scriptal misrepresentation by first defining for a given Intentional state an aboutness extension and a directedness extension; the two extensions are then compared. In this way the most common types of error can be explained: mispriming, overgeneralization, overspecification, misidentification, and errors involving mirages, optical illusions, and hallucinations. Accounting for the misrepresentation of hypothetical entities will require a different strategy. The implications of my scriptal approach to traditional externalistic and individualistic views of the mind are then examined. I concentrate on externalism, since the consequences for individualism are relatively less serious. I conclude, in chapter VIII, with a summary of what has been accomplished and a cautious look forward.

Each of the chapters to follow is meant to address one specific topic, eventually culminating in chapter VII where, hopefully, all the central issues will coherently converge. I shall do my best to express the relevance of each of the topics discussed, but where this expression is judged lacking, I must beg for the reader’s patience and indulgence. The scope and limitations of this project should be well understood at the outset, however. I do not intend my work to provide a comprehensive theory of Intentionality, but to illuminate only a specific type of Intentional mental states—vis., those states that have underlying scripts (see chapter IV). Many non-scriptal Intentional states exist, and about these I shall have nothing to say. It is important to keep this in mind, if my proposals are to be judged fairly.

It will become apparent that I owe a great deal to many who have written on the issues I take up here. But none have influenced this work more than Roger Schank, Robert Abelson, and Fred Dretske.
Intentionality

An Extended Definition

How many evils could be remedied, both on the individual and social level, by the correct psychological diagnosis, or by knowledge of the laws according to which a mental state can be modified!

Franz Brentano, Psychology from an Empirical Standpoint, p. 22.

In the previous chapter, we described Intentionality as a property of some mental states by which these states are represented as being directed toward, or about, or of something or another. Alternatively, every Intentional state was said to take something as its object. But it is not difficult to see that this description will not take us very far. Here is another offering:

Intentionality covers those characteristics of mental activities on account of which those activities are said both to have a content that contains information about something beyond the content and the activity, and to involve a particular sort of attitude towards that content. Moreover, it is a peculiarity of mental content that it is necessarily 'perspectival.' (Lyons 1995:1)

A little better, perhaps, but still not very clear. We may wonder, for instance, in what sense does an Intentional state have content containing information that goes beyond the content itself? Does this "going beyond" necessarily (regularly, occasionally) involve a real external object? And what does it mean to say that mental content is perspectival?

The literature on Intentionality contains many definitions of this sort. Each is normally followed by a lengthy explanation of the definition's various components, an explanation that is designed to bring out important elements not explicitly mentioned in the original statement. I propose something similar. It is the purpose of this chapter to provide an extended definition of Intentionality, which will afford the basis required for upcoming business. To that end, two basic distinctions and four Intentional dimensions will be introduced.

2.1 Two Distinctions

Strictly speaking, the two distinctions I have in mind are not part of the extended definition we are developing, but are meant to counter, and hopefully do away with, potential confusions and misapprehensions. The distinctions of interest are between (i) Intentional states and mental states, and (ii) Intentionality and consciousness. Let us start with the former.
1. Intentional States and Mental States

One cannot accept Brentano's thesis—that a given state is Intentional if and only if it is mental, and that mental states are distinct from physical states—unless one is also willing to accept a kind of dualism. If the statement is taken to be an expression of irreducibility, then minds are not just physical brains; they are something above and beyond. So the Brentano thesis suggests at least two fundamental sorts of things in the universe, mental and physical, which is a general statement of dualism.

If the threat of dualism is not enough to at least render suspicious, if not completely dislodge, the Brentanian conception of the mind, the following consideration might take us a little further in that direction: it turns out that not all mental states are Intentional. Some mental states have Intentionality; others do not. Searle gives the following clue to the difference:

If I tell you I have a belief or a desire, it always makes sense for you to ask, "What is it exactly that you believe?" or "What is it that you desire?"; and it won't do for me to say, "Oh I just have a belief and a desire without believing anything or desiring anything". My beliefs and desires must always be about something. (1983: 1)

In determining which mental states have Intentionality and which do not, then, we might find the following heuristic helpful. If $M$ is Intentional, there must be answers to the questions: "What is $M$ about?" "What is $M$ of or for?" "What is it an $M$ that?" "Toward what is $M$ directed?" Beliefs, desires, fears, hopes, hunches, and intuitions are Intentional in this respect, but other mental states are not so focused. "Raw feels" like pains, itches, and tickles are normally considered non-Intentional. There are also forms of elation, depression, and anxiety where one is simply elated, depressed, or anxious without being elated, depressed, or anxious about anything specific. On the other hand, one can be elated that something has occurred, or depressed and anxious at the prospect of one thing or another. On the present distinction, focused elation, depression, and anxiety are Intentional; the unfocused cases are not. Thus it appears that as far as human cognition is concerned, every Intentional state is mental, but some mental states are not Intentional. In other words, Intentional states constitute a proper subclass of mental states.

As mentioned, Brentano himself did not distinguish between Intentional and non-Intentional mental states. He believed that "feelings [including pains, itches, and tickles] undeniably refer to objects" (1874: 90); but in the case of such "raw feels," the mental state and its Intentional object—the object to which the state Intentionally refers—are somehow conjoined or united: "Even in cases where I hear a harmonious sound, the pleasure which I feel is not actually pleasure in the sound but pleasure in the hearing. In fact you could say, not incorrectly, that in a certain sense it [the auditory experience] refers to itself ... that the feeling and the object are ‘fused into one’ ... Still [mental states] retain a mental in-existence, a Subject-Object ... and the same thing is true of these feelings" (1874: 90).

Also relevant is the recent revival of representationalist views. For instance, according to Dretske's (1995) representational thesis, although not all representations are mental, all mental facts are representational facts; hence all mental facts are Intentional in one way or another. But I do not want to uphold the representational thesis in any form that renders "mental state" and "Intentional state" coextensional. Rather, I prefer the flexibility of Searle's conception of Intentionality, the view

---

1 The qualification is important because one can easily extend the notion of Intentionality to include non-intelligent systems; see Dretske's (1995) discussion of natural and conventional Intentionality. Unless otherwise stated, my discussion will concern human Intentional systems.
that traditionally takes Intentional content to be expressible by "that"-clauses. We shall later see (the
details are in §2.2.ii) that these linguistic representations often fail certain tests of substitution and
existential generalization.

ii. Intentional States and Consciousness

A further distinction must be made between Intentionality and consciousness. "Intentional
states" and "conscious states" are not coextensive, because many Intentional states are non-
conscious. The idea of non-conscious Intentionality must seem unintuitive at some level. It certainly
seemed that way to Brentano, who remarked as follows:

For any given use of the word ["consciousness"], we shall have to decide whether it may not be more harmful
than helpful. If we want to emphasize the origin of the term, doubtless we would have to restrict it to cognitive
phenomena, either to all or to some of them ... I prefer to use it as synonymous with "mental phenomenon," or
"mental act." For, in the first place, the constant use of these compound designations would be cumbersome, and
furthermore, the term "consciousness," since it refers to an object which consciousness is conscious of, seems to
be appropriate to characterize mental phenomena precisely in terms of its distinguishing characteristic, i.e., the
property of the Intentional in-existence of an object. [So] ... no mental phenomenon exists which is not, in the
sense indicated above, consciousness of an object. (1874: 102)

Brentano understood "consciousness" to mean consciousness of something, and since this "of
something" is precisely the property that the Intentional in-existence of an object is supposed to
capture, he took "conscious state" to mean exactly the same thing as "Intentional state." Moreover,
since, on his account, the mental is uniquely defined by its Intentionality, he also took "mental state"
to be coextensive with "conscious state."

We have already stipulated that some mental states are not Intentional; we can see that
Intentionality also differs from consciousness by taking belief as an example and observing that we
are not aware of all the beliefs we hold at every given moment in our lives. For instance, before now
I have never considered my belief that my grandfather lived most of his life west of the Nile Valley;
and undoubtedly I have many other beliefs, about a
multitude of different topics, that I am not
thinking about at present and have never thought about in the past. But these beliefs are just as
attributable to me when I am not thinking about them as when I am. So what we might call dormant
beliefs still constitute Intentional states in the sense that they are beliefs that such-and-such is the
case. The fact that we might not be always aware of them affects neither their content nor the fact
that they are contentful. The same reasoning applies to all the other kinds of Intentional states as
well.

The upshot is that while the class of conscious states and the class of Intentional states
overlap, they are not identical, nor is one included in the other (see Figure 2.1).

---
2 In "The Problem of Logical Omniscience, I" Stalnaker considers this to be the basic point of what he calls the Sentence
Storage Model of Belief: "To a first approximation, the idea is that to believe that P is to have a sentence that says that P
stored (to use the fashionable idiom) in one's belief box ... The explicit beliefs are those in the set of sentences stored in
the belief box." Beliefs outside the box are implicit. Fodor (1978) adheres to this picture, and assumes the sentences in
question to be literally encoded in a language of thought (see below, §3.4.i).
2.2 Four Intentional Dimensions

Now we turn to a survey of the four Intentional dimensions on which we shall focus in this project. They are: aboutness, aspectual (or perspectival) shape, directedness, and misrepresentation (or error). It is worth mentioning that these four characteristics may not be exhaustive; but aboutness, aspectual shape, directedness, and misrepresentation are surely the most important Intentional characteristics, and a reasonable account of them will take us a long way.

i. Aboutness

Many mental states have the power or capacity to refer to particular objects in the environment of the agent whose mental states they are. Eva sees Adam, hears him, has thoughts about him and desires for him. These are things Eva normally cannot do unless she occupies states that have Adam as their object, as what they refer to, as what they are thoughts about. Dretske points out that even the simplest measuring instruments exhibit this level of Intentionality:

In representing the pressure in an intake manifold, a pressure gauge "says" something about the manifold. It is not only about the manifold (an object), it is about the pressure in it (a property) and, therefore, about the manifold's having that pressure (a condition or state). If pressure gauges were conscious, if their Intentionality was original rather that conventional, the manifold... would be the object the gauge was conscious of, having a pressure of [e.g.] 14 psi would be the property the gauge perceived it to have, and its having a pressure of 14 psi would be the condition, state of affairs, or fact the gauge was aware of. (1995: 29)

We shall say that the reference of a representation is the object(s) whose properties or relations the representation designates, and the sense of a representation is the properties or relations the representation indicates the object(s) as having. The difference between represented objects and represented properties, between the reference and the sense of a representation, is what Nelson Goodman (1976) had in mind when he contrasted a picture of a black horse ("black horse" here specifying the object the picture is a representation of) and a black-horse picture ("black-horse" here specifying what the picture depicts the object as). Some pictures of black horses do not represent the

---

3 Goodman's preferred term for representational reference is "denotation."
black horse as a black horse: the picture may depict the animal as, say, a blackish spot in the distance, or perhaps the black horse had been disguised to look like a ferocious brown cow! In neither case do we have a black-horse picture: in the first instance we have a blackish-spot picture; in the second instance we have a brown-cow picture. But in both cases they are pictures of a black horse, perhaps even the same animal. Clearly, then, the sense of a representation does not determine its reference. Nor does the fact that the representation is of a given object determine how that object may be represented. So reference does not determine sense.

In *Explaining Behavior*, Dretske makes the same point by distinguishing the *comment* of a representation from the *topic* of a representation. A pressure gauge hooked to the right front tire of your car might register the pressure as 45 psi. This is the comment of the gauge’s representation: it is “telling” you that the pressure is 45 pounds per square inch. But pressure gauges do not indicate which tire pressure (if any) they are representing. If you want to know which tire pressure the gauge is representing, which *topic* the gauge is commenting on, you have to look, not at the gauge, but at the external connections between gauge and world.

In *Naturalizing the Mind*, Dretske designates this relation by “C.” What determines the topic or reference of a representation, then, is not how an object is represented, but a certain external causal or contextual relation, C, so that the object (if any) which is the reference of a given Intentional state will be that object which stands in C to the Intentional state in question.

Dan Lloyd (1989: 14) draws similar conclusions by contrasting *explicit content* and *extensional content*. Extensional content is the represented object; explicit content is the manner in which the object is represented. The represented object is not determined by the properties that object is represented as having—that is, not by the explicit content of the representation, for nothing guarantees that black-horse pictures are pictures of black horses. Many Intentional states are like pictures in this regard; they are *de re* modes of representation (Burge 1977; Recanati 1993). So

---

4 Although Dretske is critical of causal theories (especially of perception; see *Seeing and Knowing*: ch. 1), he seems to accept an account of mental-state reference (or Intentional aboutness, as I shall say), which depends on a causal relation or connection between a representation of a thing and the thing represented.

5 There is a purported difference between *de re* (“of things”) and *de dicto* (“of words”) Intentional states. *De re* beliefs are relations between agents and real objects—e.g., the belief that Anwar Sadat was president of Egypt—and are said not to be individuated solely in terms of the mental content of subjects, because the object itself (re) has to be part of the principle of individuation of the belief. *De dicto* beliefs, on the other hand, do not connect the agent with any real objects—e.g., the belief that Santa Claus comes on Christmas Eve—and can therefore be fully individuated by content alone, by what is “in the agent’s head.” It has been customary to draw the distinction in terms of a substitutivity criterion. In attributing a *de re* belief about a given object, one is free to substitute any correct description of the relevant object. If Adam believes *de re* that his car is green, then we could characterize Adam’s attitude by substituting any correct description of his car regardless of whether or not Adam would describe his car in that way. By contrast, if we say that Adam believed *de dicto* that 2 cubed equals 8, we may refuse to say that Adam believes that the cube of the *only even prime number* is 8, because Adam may not know that 2 can be so described. Burge criticizes this manner of expressing the distinction on the grounds that there are sentences where substitutivity fails at the surface level, but which are nevertheless *de re*. His alternative is to express the distinction in terms of the logical form of descriptions of belief: *purely de dicto attributions make reference to complete propositions*—entities whose truth or falsity is determined without being relative to an application or interpretation in a
looking at a faithful representation of a black horse, one is likely to believe that (s)he sees a black horse. But one can have the same type of belief, a “black-horse” belief, without it being a belief that refers to (is about) a black horse or, indeed, any object at all.

The aboutness of an Intentional state is here being understood in terms of the reference, or topic, or extensional content of that state. And if we find that some Intentional states—particularly those pertaining to imaginary or fantastic objects—do not refer to any actual thing, either directly or by means of a Kripkean causal chain, then we shall say that these states are not about anything.

If we include, as does Dretske (1995: 172-3, fn. 20) represented properties as part of what an Intentional state may be about, then an Intentional state will always be about something: when there is no object (a round square, to use Dretske’s example), the state will be about the properties (roundness and squareness) it represents the non-existent “object” as having. But we shall not follow this practice, since people not only experience and think about objects, they experience and think about their properties as well, and the two must be kept separate. If I believe that I saw Adam crossing the street one busy morning, then, if Adam was indeed in relation \( C \) to my representation of the street-crosser—i.e., if Adam was the person I saw—my belief will be about Adam. If it was not Adam I saw, if Adam was not in relation \( C \) to my representation of the street-crosser, then my belief will not be about Adam (or any Adamish properties). Rather, it will be directed toward Adam (more on directedness shortly). On the other hand, there is nothing to prevent a mental state from being about a specific property or relation if the mental state is focused on that property or relation, not on the object(s) whose property or relation it is. In this case, if the property or relation in question bears relation \( C \) to the representation, then the representation will be about that property or relation. If it does not, then the representation will not be about anything: we shall not say, for instance, that the representation is about the properties it is representing the property or relation as having.

Ultimately, then, what determines the aboutness of an Intentional state is what enters in relation \( C \) with that state. Note that \( C \) need not be direct; it can be (and often is) a long, tortuously complex causal chain mediated by physical evidence, memory, verbal or written report, etc. What is important for Intentional aboutness is that the actiology of \( C \) traces back to a real object \( a \), in which case, the relevant Intentional state will be about \( a \).

The absence of aboutness does not mean that a mental state is not Intentional, since aboutness, as it is understood here, is not a necessary component of Intentionality. To say that aboutness is not necessary for Intentionality is merely to say that one can truly believe that Santa Claus is fat and jolly, that one can have a genuine desire to meet the Tooth Fairy, that one can be really fearful of the Wolfman; and the Intentional status of such beliefs, desires, and fears is not in any way compromised on account of their aboutlessness. Nor does the claim undermine the importance of aboutness as an Intentional property. The capacity of mental states to refer to things in the environment is not just a happy twist of fate. Aboutness allows all living creatures to function at the basic level of identifying sources of food and danger, of navigating their surroundings and coping with unexpected conditions. In this respect, human are in no more need of aboutness than any other animal: all organisms must be able to find their way in a manner that is conducive to their continued existence, and no organism can do so without a robust way of representing its actual environment.

particular context. \textit{De re} locutions are about predication broadly conceived. They describe a relation between open sentences (or what they express) and objects (1977: 343). Searle (1983: ch. 8) also rejects the traditional manner of making the \textit{de re} vs \textit{de dicto} distinction, but for a different reason. He believes (contra Burge) that all Intentional states are entirely constituted by their content and their psychological mode, both of which are “in the head,” and that \textit{de re} Intentional states are only a subclass of \textit{de dicto} Intentional states.
These considerations suggest two things. First, Intentional aboutness cannot be ignored. Any account of Intentionality that omits aboutness will have failed to shed light on the connection between human cognition and the world. Second, an explanation of the aboutness dimension of Intentionality is likely to be an evolutionary one. If it is not just a coincidence that mental states have the power to refer to things outside an agent, a power that is crucial for the survival of that agent, then an account of this power is likely to be found in the evolutionary history of the agent in question. The question of why aboutness, important as it is, is not necessary for Intentionality is rather complex and must await the expository work of the next few chapters. But we can offer the following clue: the answer lies in the manner in which the brain structures whose function is to indicate the presence of external objects do their job.\(^6\)

### ii. Aspectual Shape

The Intentionality of mental states is often said to have an aspectual or perspectival dimension, because the information contained therein is necessarily slanted or from one “point of view” rather than another. Lyons explains:

If my beach bucket contains [a] football, then it contains it in a simple and straightforward way. The whole of the football, rather than one aspect of it or perspective upon it, is contained in my bucket. However, if I think about the football, only a certain perspective on or view about the football will be in my thoughts. If I am only 2 or 3 years old, then, because of the limitations of my knowledge, I may only be able to think of the football as ‘thing to be kicked that my brother got as a Christmas present’. If someone asked me whether I had just been kicking a football, I may sincerely deny that I had. For, while I had indeed been kicking a football, I was only able to think of the activity under the one and only aspect or description I knew, namely kicking the funny thing my brother got as a Christmas present. The perspectival or aspectual parameter of Intentionality is also true of all thinking, whether of an adult or a child. No matter how much I know about something, it is still limited to certain descriptions or perspectives or slants or aspects. (1995:2-3)

Dretske also acknowledges the aspectual nature of Intentionality: “In thinking about a ball,” says he, “I think about it in one way rather than another—as red not blue, as round not square, as stationary not moving. These are the aspects under which I think about the ball. I can desire an apple, yes, but in desiring an apple I desire to eat it, taste it, throw it, hold it, look at it, or simply have it. These are the aspects ... under which I desire the apple” (1995: 30-1). Most of our mental states not only have reference, aboutness, objects that constitute their topic or extensional content, but they also represent these objects in one way rather than another. Even when there is no object, there may still be an aspect: think of mirages, dreams, optical illusions, hallucinations, and so on.

Fodor speaks of the opacity of Intentional states rather than aspectual shape, but the basic point remains largely the same. He considers it a condition upon any acceptable account of the propositional attitudes that it should explain why Intentionality seems to operate under definite descriptions; he calls this “Frege’s Condition” (1978: 504). Let us say Samson believes that his pen-pal is coming to pay him a visit next week. His having this belief does not imply that—since his pen-pal is in fact a 25 year-old blond, blue-eyed woman called Delilah—he thereby believes that a 25 year-old blond, blue-eyed woman called Delilah is coming to visit, for he may not know that his pen-pal’s name is Delilah, that she is 25 years old, or that she is blond and blue-eyed. In fact, Samson may not know that his pen-pal is a woman at all. What he believes is that his pen-pal is coming to

\(^6\) See the advertisement for LASCH in §6.3.ii and the discussion of Intentional aboutness in §7.1.i.
Chapter II  

**Intentionality: An Extended Definition**

stay next week, and that is about the extent of it. Thus if someone were to tell him, “I hear Delilah is paying you a visit,” he may well deny it, and do so sincerely.

Similarly, Oedipus wanted to marry Jocasta, but his desire to do so does not entail a desire to marry his mother, even though Jocasta is his mother. In fact, Oedipus contemplated a mother-son union with all the repugnance one might naturally expect. Fodor believes that examples such as these show the opacity of our Intentional states. We can expect Oedipus’ desire to marry Jocasta to have different causal powers (including motivational force in regard to his action) and different explanatory roles (including making sense of his actions to himself) under different descriptions of his chosen bride, descriptions that may well be opaque or blind to one another.

I take the aspectual nature of Intentionality to be a rather unremarkable fact, a mere recognition of the obvious limitations to which any thinking entity is confined. Again, these limitations are by no means extraordinary, and they are not restricted to human psychology. As Lyons points out,

> every contact between any two things in the universe must be ... ‘aspectual’ or ‘from a particular point of view’. When one billiard-ball bumps into another, the first ball hits or scratches or at least bumps into just one side of the other billiard-ball. When we first catch sight of a friend in the street, we see the front or side or back of her. So it should be no surprise that when we find our way in the world or act in regard to the world or make plans in regard to the world ... then all these activities will be aspectual. So it should be no surprise that when we believe or desire something, then we do so in an aspectual way, or with a limited point of view. (1995:62)

It is precisely because of this “limited point of view” that there is a difference between Samson’s believing that his pen-pal is coming to visit and his believing that his visitor is a 25 year-old blond, blue-eyed woman named Delilah. And we know there is a difference between the two beliefs because having one does not entail having the other. Similarly, wanting to marry Jocasta does not entail wanting to marry mom, believing that a pitcher contains water does not entail believing that it contains H\textsubscript{2}O, believing that Kofi Anan is of African ancestry does not entail believing that a previously elected Secretary General of the United Nations is of African ancestry ...

We can put the same point in information-theoretic terms. If a signal carries the information that \( s \) is \( F \), it does not necessarily carry the information that \( s \) is \( G \), despite the extensional agreement of \( F \) and \( G \). Even though \( F \) and \( G \) are true of exactly the same things, the information that \( s \) is \( F \) is different from the information that \( s \) is \( G \) (Dretske 1981; Israel and Perry 1990). The extensional equivalence of \( F \) and \( G \) is a further datum that is not contained in the original signal. Dretske described this fact as the essence of informational Intentionality.

The same opacity is also manifest in language, in the sentences which describe what someone believes, desires, intends, fears, etc. In the case of language, however, we call the corresponding phenomenon intensionality (with-an-s). One of the ways a sentence can qualify as intensional is if the replacement of predicate expressions by coextensive predicate expressions alters (or can alter) the truth-value of the sentence as a whole.

So, for example, “He believes that \( s \) is \( F \)” is an intensional sentence (and the attitude or state it describes is an Intentional state) because even if \( F \) and \( G \) are coextensional (true of exactly the same things) one cannot substitute \( g \) for \( F \) in this sentence without risking a change in truth value. That is, “He believes that \( s \) is \( G \)”

---

7 Upon hearing from the oracles that he would kill his father and marry his mother, Oedipus left Corinth in the hope of avoiding this fate. Eventually, when the true nature of the marriage was discovered, Jocasta committed suicide and Oedipus blinded himself in grief and became a wandering beggar. See Sophocles’ *Oedipus the King*. 

---
may be false even though everything that is $F$ is $G$ and vice versa ... The sentences that describe what a person intends, believes, knows, hopes for, wishes, plans, and imagines are all intensional in one way or another. (Dretske 1981:75-6)

Another way a sentence may qualify as intensional is if it does not permit existential generalization as a valid form of inference. For example, from the truth of the sentence, “The sheriff believes that Mr. Holland is an honest man,” we cannot validly infer that $(\exists x) (\text{the sheriff believes that } x \text{ is honest})$, since $x$ (Mr. Holland) can just be a figment of the sheriff's imagination. On the other hand, from the truth of the non-intensional sentence, “Mr. Holland is an honest man,” we can validly infer $(\exists x) (x \text{ is honest})$ and we can substitute coextensional expressions for “Mr. Holland” without changing the truth-value of the sentence.

What all of this shows is that there are many distinct perspectives or aspects to things, and that Intentionality is sensitive to these aspects. If Dretske is correct (as I believe he is) in observing that the information that $s$ is $F$ is different from the information that $s$ is $G$ even though $F$ and $G$ are coextensional, then the aspectual nature of Intentionality can be captured by information theory. We will return to this issue in chapter VI.

iii. Directedness

We said that knowing that a certain picture is a picture of a black horse does not tell us whether it is a black-horse picture, a blackish-spot picture, a brown-cow picture, or any sort of object-picture at all. The same is true of Intentional states. Knowing that Eva has a belief $about$ Adam does not tell us what she believes, for she may have mistaken Adam for the milkman, the postman, the pizza delivery boy, or any one of many other objects we can easily imagine. In such cases we shall say that Eva’s belief is $about$ Adam, but is $directed$ toward the milkman, the postman, or the pizza delivery boy, as the case may be.

Whereas aboutness is meant to capture the topic, or reference, or extensional content of an Intentional state, the directedness of an Intentional state is meant to capture the state’s comment, or sense, or explicit content. Another way to say more or less the same thing is that aboutness is fixed widely whereas directedness is fixed narrowly. It follows that a mental state may be about $x$ (have $x$ as its wide content) but be directed toward $y$ (have $y$ as it narrow content), where $x \neq y$.

It was also said that an Intentional state need not be about anything, but that all Intentional states are directed. Having seen someone whom I believe to be Adam crossing the street, whether or not my belief is about Adam depends on whether or not Adam bears relation $C$ to my representation of the street-croosser. If Adam bears that relation, then my belief is both about Adam and directed toward him; if Adam does not bear that relation, then my belief is about whatever object bears relation $C$ to my representation, and no object means no aboutness. But regardless of whether or not anything bears $C$ to my representation, or what (if anything) that object may be, my belief is still directed toward Adam, since it is Adam whom I believe to have crossed the street.

---

8 Notice that according to the substitutivity test of determining $de re$ and $de dicto$ Intentional states, this means that the propositions expressing $de re$ Intentionality will always be non-intensional, while the propositions expressing $de dicto$ Intentionality will always be intensional; see note 5.

9 I want to keep things simple for now. More complex cases—e.g., “Samson is stronger than Adam,” and “I saw Mommy kissing Santa Claus”—will be dealt with later.
Aboutness focuses on an object exclusively: given an Intentional state $m$ that is in relation $C$ with an object $o$, for all $p$, if $m$ is about $p$, then $p$ is identical to $o$. But, though I have so far spoken of the directedness of an Intentional state as if it likewise focused on an exclusive object, directedness (unlike aboutness) is very much dependent on whether or not the agent is able to discriminate $o$ from other objects; and the ability of the agent to discriminate $o$ from other objects is a function of the kind of information the agent receives from $o$ and the manner in which this information is processed. My mental state is about Adam only if Adam stands in $C$ to my experience, and if Adam stands in relation $C$ to my experience, my mental state cannot be about any object that is not Adam. But whether or not my mental state is directed toward Adam in a similarly exclusive way depends on the resources of the information-processing system that I am, on whether or not these resources are sufficient to differentiate Adam from other superficially similar objects. If not, then my mental state will still be directed toward Adam, but not exclusively.

The intuition here is that if an information-processing system (agent) neither has the capability nor the resources to distinguish certain objects, then we should not look to the directedness of that system’s Intentional states to do the job. Rather the opposite: we should expect this kind of liberality to manifest itself in the directedness of the system’s Intentional states precisely because the system is incapable of making the relevant distinctions. The reason for this will become clear following our discussion of scripts.

Since it is possible for an Intentional state to be about $x$ but be directed toward $y$ ($x \neq y$), a question now suggests itself: How should Intentional states be individuated? If we choose to individuate Intentional states by their aboutness, the sense-reference distinction becomes irrelevant; the only thing that counts is reference. Thus the difference between believing that Delilah is coming to visit and believing that Samson’s pen-pal is coming to visit would be ignored: the two beliefs are identical because they have the same topic. Wanting to marry Jocasta is wanting to marry mother, since the reference in both cases is to the same object. And it does not matter that you think you see a brown cow and I think I see a blackish spot; our beliefs are the same because they are both about a black horse.

This is obviously inadequate. Believing that Delilah is coming to visit and believing that Samson’s pen-pal is coming to visit are not the same Intentional state, since Samson can be in one without being in the other. And even though Oedipus wanted to marry Jocasta, he certainly had no desire to marry his mother. Further, while it is true that your belief and mine are about the same black horse, it is also true that you believe you see a brown cow and I believe I see a blackish spot: the two beliefs are different in content. So if someone wanted to know what you and I believe, individuating our mental states by aboutness will be of little use. Ultimately, differences in sense make for differences in Intentional states, but differences in sense (where the reference remains constant) are impossible to capture if Intentional states are individuated by aboutness.

There is the further fact that some Intentional states have no aboutness at all. Eva believes that there is a colony of Leprechauns living inconspicuously in a forest 60 miles northwest of Dublin. She is worried that the Leprechauns will be found before she gets the opportunity to see them and (according to legend) share in their riches. How shall we treat Eva’s belief? There are no Leprechauns, no colony, and (we may suppose) no forest 60 miles northwest of Dublin. Accordingly, Eva’s belief with regard to the existence and location of the colony, her wish to see it, and her

---

10 The word "discriminate" is meant to be neutral as to the manner of discrimination.
anxiety that it may be discovered before she gets her wish are Intentional states that are not about anything. But they are Intentional states nonetheless, which need to be individuated like the rest. The same is true of Eva’s beliefs that Robinson Crusoe was lonely, that David Copperfield worked in a blacking factory, that dragons spit fire and unicorns fly. Because aboutness is not a necessary condition for Intentionality, it makes for a poor method of individuating Intentional states.

I do not mean to suggest that individuating Intentional states by aboutness is never appropriate. If I decide to organize my picture album by theme, I am more likely to do so by sorting the pictures by topic rather than by comment. A blackish-spot picture and a brown-cow picture would then be judged “the same” relative to theme or category if they were both pictures of a black horse. Similarly, if for some reason someone wanted to categorize all objects capable of producing Intentional states in humans, (s)he might choose to do so by individuating Intentional states by aboutness (which, by the way, would be the best way to proceed). So one need not deny circumstances where individuation by aboutness may be more appropriate or efficient than individuation by directedness. What should be kept in mind, however, is that the former method is much more broad-grained than the latter, that in choosing to individuate Intentional states by aboutness we risk overlooking genuine differences between these states and missing a host of others.11

Individuating Intentional states by their directedness seems like a better strategy, one that has the advantage of not requiring that the states thus individuated be about anything. So Eva’s beliefs with regard to Leprechauns, her desire to find their colony and share their riches, her fear of not doing so in time, etc. can now be recognized as full-fledged instances of Intentionality in a manner that would have been impossible if Intentional states were individuated by aboutness.

Our ability to individuate Intentional states either by aboutness or by directedness (at least in principle) suggests that we can create two extensions for any given Intentional state relative to these two dimensions. Thus Eva’s belief that Leprechauns exist has $\emptyset$ as its aboutness extension ($E_a$) and $\{l \mid l$ is a Leprechaun} as its directedness extension ($E_d$)—which is just another way of saying that Eva’s belief is not about anything but is directed toward Leprechauns. Similarly, Eva’s mistaking Adam for the milkman is a belief that has $\{a \mid a$ is Adam} as its $E_a$ and $\{m \mid m$ is the milkman} as its $E_d$.

But notice the following difference between the two cases: Eva’s belief about Leprechauns does not involve a real object, which is why it is aboutless and is directed toward an imaginary entity. Her mistaking Adam for the milkman, on the other hand, is both about a real thing (Adam) and directed toward a real thing (the milkman). The difference between the two cases raises the following concern:

The set $\{l \mid l$ is a Leprechaun} just is the empty set, since there are no Leprechauns; the identity of sets, after all, is determined by their members. So it isn’t very clear how the aboutness and directedness extensions of Eva’s Leprechauns-belief differ.12

---

11 Actually, individuating by aboutness is not always more broad-grained than individuating by directedness. We can have two Intentional states directed identically but with different aboutness (Dr. Paul Bartha has reminded me of this possibility in personal correspondence; see the discussion of “Twin-Earth Cases” in §7.1.iv). Nevertheless, my point still stands. I suggest that the number of unique Intentional states capturable by a directedness-based individuation method is significantly higher than the number of unique Intentional states capturable by an individuation method based on aboutness.

12 The point was raised by Dr. Steve Savitt in personal correspondence.
The same criticism seems to apply to any Intentional state involving a fictitious object. To answer it, we need to backtrack and take stock of what we know with respect to the Intentional status of Eva's belief.

We know that (i) Eva's belief that Leprechauns exist is an Intentional state that is somehow connected to the imaginary objects, Leprechauns; and (ii) Eva's belief that Leprechauns exist is not about anything because Leprechauns do not exist.

Now (i) and (ii) imply that whatever it is that makes Eva's mental state Intentional need not be an actual thing in the real world, that Intentionality need have no reference, or topic, or extensional content. In chapter VII I shall argue that what makes Eva's mental state Intentional is that it has directedness toward a specifiable object (not necessarily a real one) as determined by a primed definitional script (script₀). The trick is to find a way of defining the directedness extension of an Intentional state that renders irrelevant the (in)existence of its members. One way to do so is to conceive of the object(s) of directedness counterfactually, not by the items that do answer to or satisfy the definitional script in question, but by any (actual or possible) item that would satisfy the script₀.

Defining a directedness extension in this way still allows us to say definitely what does and does not go into the directedness extension of an Intentional state; so we can say that Eva's belief is directed toward Leprechauns but not toward baboons, television sets, or the third-world economy.¹³ It does not matter that Leprechauns do not exist; what matters for directedness is that an object o can exist and that o would satisfy the corresponding definitional script if it did exist. The difference between Eva's directedness extension being {l | l is a Leprechaun} and her directedness extension being Ø can now be seen as the difference between saying that the possible object Leprechaun would satisfy Eva's script₀ and saying that there is no possible object that could do so.¹⁴ In any case, directedness does not always involve abstract entities—as when Eₐ = Eₐ for some Intentional state (see §7.2.i for more details)—and so separating Eₐ and Eₐ seems well in order. Such a separation will also figure importantly in our later account of misrepresentation (more on this in chapter VII).

Returning to the main point, there is reason, then, to believe that individuation by directedness is at least more fine-grained than individuation by aboutness. Yet individuating Intentional states by directedness may not be wholly unproblematic. It is not clear, for instance, how this method can preserve the aspectual shape of Intentional content. There is no special problem with cases like Eva's mistaking Adam for the milkman, where the aboutness and directedness of a mental state have different extensions. But what about Oedipus' desire to marry Jocasta and not his mother? The milkman is a different person from Adam, not some aspect or description of him; so we may account for Eva's mistake by comparing the aboutness extension of her belief, Eₐ = {a | a is Adam}, with the directedness extension of her belief, Eₐ = {m | m is the milkman}. Oedipus' mother, however, is an aspect or description of Jocasta; so it seems that the aboutness extension of Oedipus' desire just is the directedness extension of his desire: both contain the same single object.

¹³ Notice also that we can positively deny that Eva's belief is directed toward the Seven Dwarfs, the Keebler Elves, or the Smurfs, even though all of these 'creatures' are physically similar to Leprechauns, and none of them is real.

¹⁴ Needless to say, my speaking of Leprechauns, Elves, Smurfs, and similar fictions does not commit me to the claim that such creatures exist, either here or in some other possible world. My speaking as if they do exist is only a pedagogical convenience, and I shall continue to refer to them more or less definitively in order to facilitate discussion.
Asserting that Oedipus' desire is directed toward Jocasta but not toward his mother requires an account of how directedness can focus on different aspects of the same object. This is not an extra problem with which we shall have to contend. The aspectual or perspectival parameter of Intentionality has already been acknowledged; and in showing how our scriptal approach captures this Intentional feature, we will have also shown how Oedipus's Intentional state can be directed toward Jocasta as a lover, not as a parent.

iv. Misrepresentation

Intentional misrepresentation was first pointed out by Chisholm (1957), and later stressed by Dretske (1986) and Lyons (1995). In fact, we have been talking about misrepresentation throughout this chapter. Beliefs and desires sometimes represent their Intentional objects as such-and-such when they are not such-and-such; that is, an Intentional state can "say" that $s$ is $F$ when $s$ is not $F$: Eva mistakes Adam for the milkman, her gold fish for a silver dollar, a bottle of beer for a bottle of cider; Oedipus misrepresents his mother as his lover, and his lover as not his mother; while Samson can mistake virtually anyone for his pen-pal.

The fact that Intentional states can misrepresent the world in this manner is the best support we have for the difference between Intentional aboutness and Intentional directedness. That an Intentional state can be directed toward an object that is distinct from that which stands in relation $C$ to it is the most basic and straightforward way in which error can be generated. Thus Eva’s perceptual experience of Adam, her standing in relation $C$ to him, may not depict who it is her experience is an experience of; after all, she might see him on a foggy night and believe that she sees someone else. But if Adam stands in the right causal relation to her experiences, then Eva’s belief is about Adam. Because she thinks she sees someone else, her belief is false. Echoing Goodman, Eva’s belief need not be an Adam-belief in order to be a belief about Adam.

What we are after here is simply a recognition that error goes hand in hand with Intentionality, that a representation of an object $o$ can represent $o$ as something other than it is. Having achieved that much, we defer a detailed discussion of the subject until chapter VII.

Figure 2.2 summarizes our discussion so far.

15 Notice just how difficult it would be to say this if Oedipus’ representations were purely linguistic. In this case, it would be very confusing to say that Oedipus misrepresents his mother as his lover and his lover as not his mother, since his mother and his lover are the very same object to which the name "Jocasta" rigidly refers. But Oedipus’ representations of the object whom he considers to be his mother and the object whom he considers to be his lover are very different, partly because these representations are not linguistically-based (see §4.4.ii), and partly because they do not designate rigidly (see §7.1.iii).
Chapter II  Intentionality: An Extended Definition

Figure 2.2. A summary diagram stressing some of the points covered in this chapter. Intentional states are differentiated from mental states and conscious states. Four dimensions or characteristics of Intentionality are also represented: aboutness, aspectual (or perspectival) shape, directedness, and misrepresentation (or error).

Intentional States

- **Aboutness**: the reference, topic, or extensional content of an Intentional state; whatever stands in relation $C$ to a state with Intentionality.

- **Aspectual Shape**: Intentional states impart information that is slanted—i.e., from one perspective rather than another.

- **Directedness**: the sense, comment, or explicit content of an Intentional state. An Intentional state may be about an object, $x$, but be directed toward an object, $y$ (either real or invented), where $x \neq y$. Although Intentionality may be aboutless, no Intentional state is undirected.

- **Misrepresentation**: Intentional states can indicate that $x$ is $F$ when $x$ is not $F$. Misrepresentation constitutes fairly strong evidence for the aboutness-directedness distinction.

Two Distinctions

- **Intentional States vs. Mental States**
- **Intentionality vs. Consciousness**

The class of mental states encompasses the class of conscious states and the class of Intentional states. Area 1 represents conscious mental states that are not Intentional, e.g., the unfocussed elation and depression mentioned in §2.1.i. Area 2 represents those mental states that are both conscious and Intentional—e.g., explicit beliefs and desires. Area 3 represents non-conscious mental states that are Intentional—e.g., the “dormant” beliefs discussed in §2.1.ii. Area 4 represents the class of mental states that are neither conscious nor Intentional, such as an agoraphobia-related general anxiety that a victim mistakes for a work-induced bad mood.
Skin, Boundaries and Authority

Locating Intentional Content

By 'intuition' I do not mean the fluctuating testimony of the senses or the deceptive judgement of the imagination as it botches things together, but the conception of a clear and attentive mind, which is so easy and distinct that there can be no room for doubt about what we are understanding. Alternatively, and this comes to the same thing, intuition is the indubitable conception of a clear and attentive mind which proceeds solely from the light of reason.


How can I tell what I think till I see what I say?


3.1 Background to Externalism

The internalist-externalist controversy is really a dispute about psychological boundaries and authority, and it goes back at least to Descartes. Descartes believed that we know some of our propositional mental events in a direct, authoritative, non-empirical manner, and that the paradigm of this self-knowledge is the *cogito*. He famously held that the occurrence of thought guarantees the existence of a thinker. A version of this insight appears in the canonical slogan, *cogito ergo sum: I am thinking, therefore I exist* (Discourse IV: 127). As illustrated early in Meditations II, the purported insight has it that while the existence of a body may be subject to scepticism, the existence of a person—qua *thinker*—looks to withstand even the most hyperbolic doubts. Thus Descartes confidently proclaimed: though the Evil Genius may "deceive me as much as he can, he will never bring it about that I am nothing so long as I think that I am something" (Meditations II: 17)—i.e., the very attempt to doubt one's own existence is paradoxical, since every such effort *is* an occurrence of thought and the occurrence of thought requires a thinker.

Descartes maintained that since we can know what we think authoritatively, the nature of our thoughts must be independent of the material world. A parallel inference is proposed in his distinction between mind and body in Meditations VI, and in Principles of Philosophy (I: 60):

---

1 All page references of Descartes' work are to Cottingham, Toothoff, and Murdoch's (1984) two-volume set, *The Philosophical Writings of Descartes*. The slogan "I am thinking, therefore I exist" has inspired a number of classroom witticisms, of which Lyons (2001: 6) mentions one: "You can't, of course, substitute 'I am, therefore I think', ... for that would be to put Descartes before the horse." See Guirgis (2002) for a review of Lyons' excellent book, *Matters of the Mind*. 

31
... from the mere fact that each of us understands himself to be a thinking thing and is capable, in thought, of excluding from himself every other substance, whether thinking or extended, it is certain that each of us ... is really distinct from every other thinking substance and from every corporeal substance.2

Descartes observed that he had "clear and distinct ideas" of his thoughts, and then reasoned that since he can separate thought from corporeal existence, the process of thinking must be independent of his physical environment and of other beings. Tyler Burge suggests that "this line of argument guarantees the truth of individualism" (1988: 113).

Putting individualism aside for the moment, the troubles with Cartesian dualism are formidable and all too familiar. In the fifth set of objection to the Meditations, Pierre Gassendi asks: "How can there be effort directed against anything, or motion set up in it, unless there is mutual contact between what moves and what is moving? And how can there be contact without a body... ?" (Objections and Replies: 237). Writing to Descartes in 1643, Princess Elizabeth of Bohemia expresses a similar concern. She requests an explanation of "how man’s soul, being only a thinking substance, can determine animal spirits so as to cause voluntary action" (Wilson 1969: 373). At her urging, Descartes published The Passions of the Soul in 1649 to clarify these matters. But his reply—that the body causes the soul to have passions, and the soul causes the body to move through an inexplicable causal union—did not satisfy the Princess. Descartes' conception of the mental and physical as metaphysically separate invited difficult problems which he seemed unable to resolve.4

A former professor of mine once remarked that to call someone a dualist now is considered an insult, that Cartesian dualism—with all its attributions of privileged access, incorrigibility, infallibility, etc.—has hindered real progress in the philosophy of mind for decades. I doubt both statements: the first seems unduly excessive, while the second ignores important facts about the evolution of philosophy as a process and the role of the dualist program in shaping how we currently think about and investigate the mental.

To be sure, there are few living dualists today. The confrontation between dualism and standard physics has been endlessly discussed ever since Descartes' own time, and is widely

---

2 See also Rules for the Direction of the Mind, 8: 29-30.

3 Descartes apparently thought very highly of Princess Elizabeth, enough to dedicate the Principles of Philosophy to her. In his dedicatory letter, Descartes praises her "incomparable" powers of understanding, and declares her to be the only person to have completely understood all his published works. "And when I consider that such a varied and complex knowledge of all things is to be found not in some aged pedant who has spent many years in contemplation but in a young princess whose beauty and youth call to mind one of the Graces rather than gray-eyed Minerva or any of the Muses, then I cannot but be lost in admiration" (Principles of Philosophy, Dedicatory Letter: 192).

4 One of the most blunt criticisms of Descartes—the "knot"—came from "a group of philosophers and geometers" at the end of the sixth set of objections: "The knot is this. We perceive very well that three and two make five and that if you take equals from equals the remainders will be equal; we are convinced of this and numerous other matters, just as you find yourself to be. But why are we not similarly convinced on the basis of your ideas, or our own, that the soul of man is distinct from the body, or that God exists? You will say that you cannot graft this truth into us unless we are prepared to meditate along with you. Well, we have read what you have written seven times, and have lifted up our minds, as best we could, to the level of the angels, but we are still not convinced. We do not believe you will allege that our minds are in a grip of a brutish stupor and are wholly unfitted for metaphysical subjects, when we have had thirty years practice in them! Surely you will prefer to accept that your arguments derived from the ideas of the mind and of God do not have the kind of weight or strength that could or should conquer the minds of learned men ... " (Objections and Replies: 283-4). But by far the most heated exchange—containing numerous ad hominem attacks and personal insults—was between Descartes and the Jesuit, Pierre Bourdin; see the seventh set of Objections and Replies and Descartes' Letter to Father Dinet.
regarded as the inescapable and fatal flaw of any dualistic approach to cognition (Dennett 1991: 35). Nonetheless, it does not require much imagination or a great deal of philosophical acumen to endlessly pick on, criticize, disparage, belittle, blame, or plain bad-mouth an admittedly problematic view without taking the time to think about its place in intellectual history. At the very least, we can credit Cartesianism—as a clear expression of a longstanding model of the mind—with having provoked the critical analysis requisite for theoretical progress; that is how philosophy evolves.

Even though dualism has been largely abandoned, an offspring sentiment still lingers, an ideal that likens the relation between a person and his thoughts to seeing, where vision is taken to be a kind of direct, immediate experience. On the most unqualified versions of this picture, a person’s inspection of the contents of his thoughts is infallible; the notion of incompletely understanding them has no application whatsoever. This model fastens on to the facts that we are generally good at identifying a wide variety of our own Intentional states, and that we have at least a prima facie authority in reporting them.

The notion of our having direct or privileged access to some of our thoughts has been variously expressed, invoked, or presupposed in the theorizing of many contemporary philosophers. For instance, it seems to be implicitly operative in Frege’s distinction between the sense of a name and the reference of a name. The following passage, taken from a letter Frege wrote to Philip Jourdain in 1914, is striking.

Let us suppose an explorer travelling in an unexplored country sees a high snow-capped mountain on the northern horizon. By making enquiries among the natives he learns that its name is ‘Alpha’. By sighting it from different points he determines its position as exactly as possible, enters it in a map, and writes it in his diary: ‘Alpha is at least 5000 meters high’. Another explorer sees a snow-capped mountain on the southern horizon and learns that it is called Ateb. He enters it in his map under this name. Later comparison shows that both explorers saw the same mountain. Now the content of the proposition ‘Ateb is Alpha’ is far from being a mere consequence of the principle of identity, but contains a valuable piece of geographical knowledge. What is stated in the proposition ‘Ateb is Alpha’ is certainly not the same thing as the content of the proposition ‘Ateb is Ateb’. Now if what corresponded to the name ‘Alpha’ as part of the thought was the meaning [i.e., referent] of the name and hence the mountain itself, then this would be the same in both thoughts. The thought expressed in the proposition ‘Ateb is Alpha’ would have to coincide with the one in ‘Ateb is Ateb’, which is far from being the case. What corresponds to the name ‘Ateb’ as part of the thought must therefore be different from what corresponds to the name ‘Alpha’ as part of the thought ... An object can be determined in different ways, and every one of these ways of determining it can give rise to a special name, and these different names then have different senses; for

---

5 Ryle (1949: 12-13) puts the matter this way: “Even when ‘inner’ and ‘outer’ are construed as metaphors, the problem how a person’s mind and body influence one another is notoriously charged with theoretical difficulties. What the mind wills, the legs, arms and the tongue execute; what affects the ear and the eye has nothing to do with what the mind perceives; grimaces and smiles betray the mind’s moods and bodily castigations lead, it is hoped, to moral improvement. But the actual transactions between the episodes of the private history and those of the public history remain mysterious ... They could not be reported among the happenings described in a person’s autobiography of his inner life, but nor could they be reported among those described in someone else’s biography of that person’s overt career. They can be inspected neither by introspection nor by laboratory experiment. They are theoretical shuttlecocks which are forever being bandied from the physiologist back to the psychologist and from the psychologist back to the physiologist.” Of course, mind-body interaction is also a problem for materialism. The difference, however, is that the materialist expects questions related to mental causal efficacy to disappear once we have a mature physical theory and/or cognitive science. The same cannot be said of substance dualism, which seems to generate difficulties that are unsolvable in principle (unless, that is, there is a wholesale revision and significant expansion of what we call “physics,” which is now limited only to the material universe).

6 Beyond philosophy, Descartes’ “two world” view had important ramifications for the development of psychology. Not only did it enabled psychologists to claim that they had a science of their own, quite independent of physiology, but it also stirred psychologists to create unique investigative methods especially tailored to their newly discovered discipline (Lyons 2001: 18-19).
Chapter III  Skin, Boundaries, and Authority

it is not self-evident that it is the same object which is being determined in different ways. (Philosophical and Mathematical Correspondence: 80)

In other words, someone who hears and understands an utterance of a sentence containing the singular term “Alpha”—such as “Alpha is at least 5000 meters high”—must think of the mountain. But, as Dretske remarks, it is not possible to think of a mountain (or a ball, or an apple, or anything else) save in a particular way (1995: 30-1). If one gives the way in which our subject was dunking of the mountain, one would be giving what Frege calls the sense the subject attaches to the name “Alpha.” It is that sense, or aspectual shape, to which the subject has direct and privileged access.

Russell also makes explicit use of the idea of direct access:

Whenever a relation of supposing or judging occurs, the terms to which the supposing or judging mind is related by the relation of supposing or judging must be terms with which the mind in question is acquainted. This is merely to say that we cannot make a judgement or a supposition without knowing what it is that we are making our judgement or supposition about. It seems to me that the truth of this principle is evident as soon as the principle is understood. (1917: 160)

We have acquaintance in sensation with the data of our outer senses, and in introspection with the data of what may be called the inner sense—thoughts, feelings, desires, etc.; we have acquaintance in memory with things which have been data either of the outer senses or of the inner sense. Further, it is probable ... that we have acquaintance with Self, as that which is aware of things or has desires towards things. (1912: 28)

For Russell, acquaintance is direct, infallible, and non-propositional; and because it is non-propositional, being acquainted with x does not imply having knowledge about x, since Russell believed that knowledge about is essentially propositional in nature. Moreover, because acquaintance is direct, it does not come piecemeal; it is an all or none achievement.

Along with logical positivism’s effort to reduce all folk psychological idioms to statements couched in the language of physics, last century’s most conspicuous attempt to dispel the notion of direct Intentional access came from the behaviourist faction. Most forms of behaviourism claim that mentalistic attributions can be analytically defined, or given strict meaning equivalences, purely on non-mental grounds. Gilbert Ryle, for instance, stoutly maintained that there is no distinction

---

7 See the first chapter of Evans (1982) for a good introduction to Frege’s sense-reference (or sense-meaning) distinction.

8 Frege also wrote the following to Russell in 1902: “The thought that all thoughts belong to class M are true is different from the thought that all thoughts belong to class N are true, for someone who did not know that M coincides with N could hold one of these thoughts to be true and the other to be false” (Philosophical and Mathematical Correspondence: 153). The same idea is being expressed here once more: even though class M and class N are identical, the thought that all thoughts belong to class M are true is different from the thought that all thoughts belong to class N are true, and for exactly the same reason that the thought Alpha is Ateb is different from the thought Alpha is Alpha: the senses of the two thoughts are not the same—which is why only one of the propositions about the mountain “contains a valuable piece of geographical knowledge.” Compare this to Dretske’s (1981) claim that if a signal carries the information that s is F, it does not necessarily carry the information that s is G despite the extensional agreement of F and G.

9 “Acquaintance,” says Russell, “which is what we derive from sense, does not, theoretically at least, imply even the smallest ‘knowledge about,’’ i.e. it does not imply knowledge of any proposition concerning the object with which we are acquainted. It is a mistake to speak as if acquaintance had degrees: there is merely acquaintance and non-acquaintance. When we speak of ‘becoming better acquainted,’ as for instance with a person, what we must mean is, becoming acquainted with more parts of a certain whole; but the acquaintance with each part is either complete or non-existent. Thus it is a mistake to say that if we were perfectly acquainted with an object we should know all about it. ‘Knowledge about’ is knowledge of propositions, which is not involved necessarily in acquaintance with the constituents of the propositions” (1914: 151).
between first-person and third-person access to Intentional states. In both cases the process is essentially the same: ordinary inspection of workaday behaviour gives rise to the discovery of patterns which lead to the imputation of appropriate propositional attitudes. Hence we know our own minds in exactly the same way we know the minds of others: by observing what we do and what we say (1949: 155).

B. F. Skinner, to take a more extreme example, had a passionate disdain for Intentional descriptions, especially those positing a "little man or homunculus":

Primitive origins are not necessarily to be held against an explanatory principle, but the little man is still with us in relatively primitive form. He was recently the hero of a television program called "Gateways to the Mind," one of a series of educational films sponsored by the Bell Telephone Laboratories and written with the help of a distinguished panel of scientists. The viewer learned, from animated cartoons, that when a man's finger is pricked, electrical impulses resembling flashes of lightning run up the different nerves and appear on a television screen in the brain. The little man wakes up, sees the flashing screen, reaches out, and pulls a lever. More flashes of lightning go down the nerves to the muscles, which then contract, as the finger is pulled away from the threatening stimulus. The behavior of the homunculus was, of course, not explained. An explanation would presumably require another film. And it, in turn, another. (1964:80)

Positing homunculi to aid psychological explanations need not be objectionable, just as long as those "little men" are properly discharged (Dennett 1991: 14). Skinner knew this. His objection to folk psychology was in fact more profound than the above criticism suggests. To put it bluntly, the folk psychological way of talking about and explaining human behaviour was for him utter nonsense; and so was any attempt to analyze human behaviour using the categories of folk psychology.

What Skinner despised most of all was the view of the mind that assigns a certain measure of privacy to Intentional states, the appeal to which from a third party was consequently assumed to be inferential. In particular, internalism was decried, for it has "the effect of diverting attention from the external environment" (1971: 195). Skinner's strategy for dealing with the "privacy" of thought was to advise that an adequate science of behaviour must consider events taking place within the skin of the organism, not as physiological mediators of behaviour, but as part of behaviour itself. A science of behaviour can deal with these events without assuming that they have any special nature or must be known in any special way. "The skin is not that important as boundary," says he, "private and public events have the same kinds of physical dimensions" (1964: 84). He based his explanation of human behaviour on the simple (and somewhat simplistic) reinforcement of operants model:

We change the relative strengths of responses by differential reinforcement of alternative courses of action; we do not change something called a preference. We change the probability of an act by changing a condition of deprivation or aversive stimulation; we do not change a need. We reinforce behavior in particular ways; we do not give a person a purpose or an intention. (1971: 94; my emphasis)

This passage suggests that there is no place for "consciousness" or "awareness" in a science of human behaviour, and as far as Skinner was concerned, there wasn't. He argued that an organism learns to react discriminatively to the world around it under certain contingencies of reinforcement. "Thus a child learns to name a color correctly when a given response is reinforced in the presence of the color and extinguished in its absence ... So far as we know, the same process of differential reinforcement is required if the child is to distinguish among the events occurring within his own skin" (1964: 85). This makes "consciousness" and "awareness" social products: "We learn to see that we are seeing only because a verbal community arranges for us to do so" (1964: 88). There are some surprising ramifications of this position: because the verbal community cannot reinforce self-descriptive responses as easily or effectively as it can teach a child to call one pattern of stimuli
“red” and another “orange,” Skinner concludes that “a person cannot describe or otherwise ‘know’ events occurring within his own skin as subtly or precisely as he knows events in the world at large” (1964: 85).

Skinner published “Behaviorism at Fifty” in order to commemorate the 50th anniversary of the behaviourist movement and to provide a “restatement of radical behaviourism.” The paper was read to an audience, some of whom had questions after the presentation. One listener asked,

What is Professor Skinner’s behaviouristic account of the high intellectual activities reflected in his paper?

Skinner’s answer:

May I say first that psychology needs to learn one very important lesson, and that is that it cannot answer every question which is asked of it. We cannot, now, give a very adequate account of intellectual activity—probably the most complex behavior of the most complex organism which has yet appeared on the face of the earth. (1964: 99)

One gets the sense that behaviourists have learned the lesson to which Skinner refers better than anyone else; for there has to be something seriously amiss with a theory that, after fifty years of growth and fine-tuning, has still virtually nothing to say about higher cognition. Though psychology is still far from giving an adequate account of human intellectual activity, a lot of progress has been made in this area over the past two decades; and the progress was made despite behaviourism, not because of it.

Behaviourism in general—and the Skinnerian brand, in particular—is now as defunct as Cartesian dualism, in part because (and here comes an amusing piece of irony) it could not explain one of the most conspicuous aspects of mentality: the fact that contentful states are generally known to the subject who has them without appeal to behaviouristic or otherwise external evidence (Davidson 1987: 95; Boghossian 1989: 152). The trouble is not merely that behaviourism runs counter to all the relevant appearance; the trouble is that, for much of what we do know about our thoughts, behaviourism can offer no explanation at all. I think I am thirsty and I know, immediately upon having this thought, that I am thirsty is what is going through my mind. Skinner’s reinforcement model has nothing to offer here, and Ryle’s denial of a difference between first-person and third-person access to mental states seems simply wrong. My knowledge of my occurrent thought could not have been inferred from any premises about my behaviour, because it could not yet have had any traction on my behaviour. Appealing to my verbal community—especially in light of Skinner’s admission that society cannot properly reinforce private events—seems flimsy at best. Sometimes I feel thirsty instantly upon waking up, and sometimes I wake up because I feel thirsty; in both cases I know directly what I am thinking. Having been sleeping for a number of hours prior, there seems to be no behavioural facts upon which my self-knowledge can be based.11

10 See Dennett’s “Skinner Skinned” (1978: ch. 4), and Davidson’s “Mental Events” (1970: 216-17).

11 Two works which should be cited here are Charles Siewert’s The Significance of Conscious Experience and Roger Moran’s “The Authority of Self-Consciousness.” The first is an appeal to have first-person phenomenal experience occupy a central place in our conception of mind and Intentionality, and the second does the same for first-person warrant and authority. See also Dretske’s (2001) comments on Siewert’s book.
3.2 The Role of the Environment

It is reasonable to construe behaviourism as a kind of anti-individualism, since it sought to identify mental states using external criteria. But though behaviourism is no longer in vogue, externalism still has a wide following. The role that behaviour was supposed to play in identifying Intentional states has now been taken over by "the environment." Externalism implies that Intentional content is not fixed exclusively by what is going on inside someone's skin or by what is accessible to a person by reflection. For the externalist, mentalistic attribution does not presuppose that the subject has fully mastered—or even is capable of fully mastering—the content of his or her mental states. In fact, since the environment is said to play a role in determining the Intentionality of the mental and no one can know everything about the environment, the externalist expects that there will be cases where the subject will not know what (s)he is thinking.

Just what role the environment plays in determining Intentional content depends on what one means by "environment." Different intuitions have produced different arguments.

i. Physical Furniture: Adam and Adam_{te}

One way to think of the environment is in terms of the things or objects that surround us. In the context of externalism, this means that the physical furniture of our lives contributes to the identity of our Intentional states. This is brought out nicely in Putnam's (1975) now famous thought-experiment, in which we are to conceive of a near duplicate of our planet Earth, called "Twin-Earth." Except for certain features about to be noted, Twin-Earth resembles Earth in every detail. The physical environments look and largely are the same. Many of the inhabitants of one planet have counterparts on the other, with identical microphysical, phenomenal, and dispositional states and histories. It just so happens that on Twin-Earth there is no H\textsubscript{2}O. The liquid that runs in rivers on the twin planet, that fills bathtubs and falls from the sky looks, tastes, and feels like H\textsubscript{2}O, but is in fact a different compound with a very different chemical structure, XYZ. The inhabitants of Twin-Earth call XYZ "water," but twin-water (water\textsubscript{te}) is not water: water is H\textsubscript{2}O.\textsuperscript{12} The year is 1750, when no one on Earth is yet aware of the molecular composition of water, and scientists on Twin-Earth have not yet discovered that water\textsubscript{te} is XYZ.

We now suppose that Adam is an English-speaking native of Earth and that Adam\textsubscript{te} is his physiological duplicate on Twin-Earth.\textsuperscript{13} When Adam and his doppelgänger simultaneously form beliefs that they express by saying, "There is water in the pitcher," what they say is different, since their respective utterances have different truth-conditions. The twins' beliefs will also be different, since their beliefs pick out different objects in their respective environments: Adam's belief picks out...
water, H$_2$O; Adam$_e$'s picks out XYZ, twin-water. So the Moral of Putnam's story is that the physiological identity of the twins does not guarantee the identity of their propositional attitudes.

The Twin-Earth story was meant to show the incompatibility of two different mind-sets. Putnam holds that many philosophers have wrongly assumed that psychological states like belief, desire, and knowing the meaning of a word are both (i) "inner" in the sense that they do not presuppose the existence of any thing or any one other than the subject to whom they are ascribed, and (ii) that these are the very states we normally identify and individuate and call the propositional attitudes. Since, as the Twin-Earth story shows, the individuation of propositional attitudes and meanings must involve relations to objects and events other than the individual, Putnam believes (i) and (ii) come apart. In his opinion, no state can satisfy both conditions. He calls psychological states satisfying condition (i) narrow, and those properly called "propositional attitudes" (i.e., those that are sensitive to agent-environment relations), wide. So a believer is not necessarily the best judge of what his own belief is about, because what a belief is about is what it "maps" or "hooks on to" in the world, facts of which the believer could be ignorant. "It is a bit like a fisherman casting his line into the river, successfully, and thinking he has hooked a trout when it is only a pike" (Lyons 1995: 84).

To put the matter another way, Adam sees a pitcher of H$_2$O and says to himself, "Here's a pitcher of water;" back on Twin-Earth, Adam$_e$ utters the same words when he sees a pitcher full of XYZ. Each speaks the truth, since their words mean different things, and since both are sincere, it is natural to suppose that each believes different things: Adam believes that there is a pitcher of water in front of him; his twin believes that it is a pitcher of water$_e$ he sees. But do they know what they believe? On the one hand, we seem to have a case where the twins are in identical narrow psychological states, but on the other, if the meanings of their words—and thus the propositional attitudes expressed by using these words—are partly determined by external factors about which the agents are ignorant, their propositional attitudes and meanings are not narrow in Putnam's sense. There is therefore nothing on the basis of which either speaker can tell which state he is in, for there is no internal or external clue to the difference. We ought, it seems, to conclude that neither speaker knows what he means or thinks.

This conclusion has been drawn explicitly by Putnam, who declares that he "totally abandons the idea that if there is a difference in meaning ... there must be a difference in our concepts (in our psychological states)" (1975: 164-5). On Putnam's view, part of what makes it true that Adam has thoughts involving water is that it is typically in re H$_2$O that he tokens those thoughts; Adam$_e$, who grew up on Twin-Earth, would not have the concept water but some other concept, water$_e$. Two individuals can therefore be in all relevant physical respects the same, and yet mean different things by their words and have different propositional attitudes (as these are normally identified).
Those who are convinced of the external dimension of Intentional content as ordinarily identified and individuated have reacted in different ways. One response has been to make a distinction between the contents of thought as subjectively and internally determined, and ordinary beliefs, desires, and intentions as we normally attribute them on the basis of outward signals and connections. Typically, this position regards the physical state of the agent as supervenient upon the agent’s narrow psychological state. It seems to me that this is the trend of Putnam’s argument.

### ii. Sociolinguistic Conventions: “Arthritis” and “Tharthritis”

The social and linguistic practices of a community are as much a part of one’s environment as any physical object. So if we allow the “environment” a foothold in shaping one’s mental states, it would appear that sociolinguistic conventions have an equal claim to define this role as do the things around us. Adopting this interpretation, Tyler Burge (1979) first argues that Intentional states are typically specified by subordinate sentential clauses, that-clauses, which may be judged as true or false; and that in an ordinary sense, the noun phrases that embed sentential expressions in mentalistic idioms provide the content of the mental state or event in question. Accordingly, he calls that-clauses and their variants, “content clauses.” Burge then offers a three-step thought-experiment, with the usual externalist conclusions.

In the first step, we are to suppose that a given person has a large number of attitudes commonly attributed with content clauses containing “arthritis” in oblique occurrence. For example, he correctly thinks that he has suffered from arthritis for years, that his arthritis seems most painful in his knees, that certain aches are characteristic of arthritis, that arthritis comes in different types, etc. In addition to these attitudes, he thinks, incorrectly, that he has developed arthritis in his thigh (arthritis is exclusively an inflammation of the joints).

The second step consists of a counterfactual supposition involving a situation in which the patient proceeds from birth through the same course of physical events that he actually does. He goes through the same motions, engages in the same behaviour, has the same sensory intake and dispositions to respond to stimuli, and participates in the same interactions with linguistic expressions. But whereas in the actual case “arthritis,” as used by the patient’s community, does not extend beyond joint ailments, in the counterfactual case the supposition is that “arthritis,” as used by the patient’s community, does include and encompass his misuse. In short, the counterfactuality in step two touches on the patient’s social environment: he might have had the same physical history and non-Intentional mental phenomena, while the word “arthritis” was conventionally applied, and defined to apply, to rheumatoid disorders that include the one in his thigh.

Burge presents the final step as an interpretation of step two. According to him, the patient in the counterfactual situation lacks all of the attitudes, with content clauses containing “arthritis” in oblique occurrences, commonly attributed to the patient in the actual situation: he lacks the occurrent thoughts or beliefs that he has arthritis in the thigh, that he has had arthritis for years, that stiffening joints and aches with a specific character are symptoms of arthritis, and so on. This is because when, in the counterfactual case, the patient utters “One can develop arthritis in the thigh,” the sentence is true, and the belief the patient expresses by his utterance is correct. But in the imagined case, neither the sentence nor the corresponding belief is correct. Switching between one linguistic community

---

18Burge’s use of “oblique” is the same as our use of “intensional” in §2.2.ii. Thus “water” is oblique (or intensional) when it functions in a content clause so that it is not freely exchangeable with extensionally equivalent terms without threat of changing the truth-value of the expression in which it appears.
and the other, the patient’s belief would change accordingly, since the meaning of the *content clause* in the relevant sentence expressing his belief would likewise change. Because the patient would have no reason to suppose that the content of his belief changed, he could not be said to have known what he believed in the first place.

The differences seem to stem from differences "outside" the patient considered as an isolated social organism, causal mechanism, or seat of consciousness. The difference in his mental contents is attributable to differences in his social environment. In sum, the patient’s internal qualitative experiences, his physiological states and events, his behaviourally described stimuli and responses, his dispositions to behave, and whatever sequences of states (non-Intentionally described) mediated his input and output—all these remain constant, while his attitude contents differ, even in the extensions of counterpart notions ... [S]uch differences are ordinarily taken to spell differences in mental states and events. (Burge 1979: 28-9)

Burge maintains that part of what makes it true that some of my thoughts involve the concept *arthritis* is that I live in, and defer to, a community in which that concept is used in a certain way. Were I to have grown in an identical community, except that in it “arthritis” covers all rheumatoid ailments, I would not have the concept *arthritis* but some other concept, *thrthritis*.

Putnam has reached similar conclusions, but, as we have seen, Putnam’s reasoning does not depend on the idea that social linguistic usage dictates (under more or less standard conditions) what speakers mean by their words, nor, of course, what their narrow psychological states are. On the other hand, not everyone has been persuaded that there is an intelligible distinction to be drawn between narrow psychological states and psychological states identified in terms of external facts (social or otherwise). Burge denies that there are, in any interesting sense, psychological states satisfying Putnam’s condition (i), in part because Burge assigns a certain primacy to language over thought. Thus, he thinks it profitable “to see the language of content attribution as constituting a complex standard by reference to which the subject’s mental states and events are estimated, or an abstract grid on which they are plotted. Different people may vary widely in the degree to which they master the elements and relations within the standard, even as it applies to them all” (1979: 79).

### iii. Causal History: Swampman

So far we have considered the environment as understood in terms of physical furniture and sociolinguistic habits. But the environment does not just include things and conventions; it includes the causal history or aetiology of events. And depending on how this history unfolds, an externalist may offer reasons purporting to show that causal-historical facts can function to define Intentional content.

But the word “history” is another variable here, for it can signify different durations, ranging from the entire evolutionary development of a species (see the following subsection) to the experiential lifespan of a single organism; in fact, anti-individualist arguments have been put forth based on both extremes. In “Knowing One’s Own Mind,” Davidson assumes a relatively local time-frame, and argues—with the aid of yet another (famous) thought-experiment—that the Intentional content of thought depends first and foremost on the causal context in which thought is produced in an agent. Little or no causal context means little or no Intentional content:

Suppose lightening strikes a dead tree in a swamp; I am standing nearby. My body is reduced to its elements, while entirely by coincidence (and out of different molecules) the tree is turned into my physical replica. My replica, The Swampman, moves exactly as I did; according to its nature it departs the swamp, encounters and
seems to recognize my friends, and appears to return their greetings in English. It moves into my house and
seems to write articles on radical interpretation. No one can tell the difference.

But there is a difference. My replica can't recognize my friends; it can't recognize anything, since it never
cognized anything in the first place. It can't know my friends' names (though of course it seems to), it can't
remember my house. It can't mean what I do by the word 'horse', for example, since the sound 'horse' it makes
was not learned in a context that would give it the right meaning—or any meaning at all. Indeed, I don't see how
my replica can be said to mean anything by the sounds it makes, nor to have my thoughts. (Davidson 1987: 91)

Davidson agrees with Putnam and Burge that the Intentional content of ordinary
propositional attitudes cannot be accounted for in terms of physical, phenomenal, causal-functional,
computational, or syntactical states or processes specified non-Intentionally and defined purely on
individuals in isolation from their physical and social environment. For Davidson, the issue depends
simply on how the basic connection between thoughts and things is established. He holds that it is
established by interactions between people and parts or aspects of the world. Dispositions to react
differently to objects and events are therefore central to the correct interpretation of a person’s
thoughts and speech; otherwise we would have no way of discovering what others think, or what
they mean by their words. “The principle is as simple and obvious as this,” says Davidson,

a sentence someone is inspired (caused) to hold true by and only by sightings of the moon is apt to mean
something like ‘There’s the moon’; the thought expressed is apt to be that the moon is there; the thought inspired
by and only by sightings of the moon is apt to be the thought that the moon is there. (1987: 100)

Davidson does not claim that all words and sentences are directly conditioned by what they
are about; we can learn to use the word “moon,” for instance, without ever seeing the actual celestial
body. The claim, rather, is that all words and thoughts must have a foundation in historical causal
connections, and these connections constrain the Intentional contents of thought and language. On
this account, Swampman’s problem becomes obvious: he has no causal history to speak of, which
means that he—at least at the moment of his creation and presumably for some indeterminate time
thereafter—has no Intentional states at all. Familiar externalist conclusions are then easily drawn:
“we are ... free,” says Davidson, “to hold that people can be in all relevant physical respects
identical while differing psychologically” (1987: 104).


In the mid 1980s, a new theory emerged that was motivated in part by the belief that the
manner in which the environment influences Intentional thought reaches far beyond the physical
furniture of our surroundings, sociolinguistic customs, or local event history. This position—which
seems to interpret “environment” very broadly to include the entire evolutionary heritage of a
species—is associated with Colin McGinn, but more so with the work of Ruth Garrett Millikan in
her book, Language, Thought, and Other Biological Categories, and subsequently in a series of
collected articles under the title, White Queen Psychology and Other Essays for Alice.

For Millikan, Intentionality is a real, non-linguistic, biological phenomenon. It has been
produced in humans in exactly the same way in which hearts, lungs, and kidneys have been produced
(hence the subtitle of her 1984 book: “New Foundations for Realism”)—that is, as a tool that
performs specific tasks. Just as any tool may be used in ways other than those for which it was
designed, or may in time lose the ability to perform its function adequately, so too a bodily organ

19 See the Reviews by Jay Rosenberg (1987) and Peter Godfrey-Smith (1988).
may, through misuse or damage, fail to perform the tasks it was designed to carry out. In general, we must look to the effects a tool or a device is designed to have in order to understand its proper function. In the case of both human and non-human organisms, the design in question will be that which has been brought about by evolution through the process of natural selection. Millikan suggests that we should look upon our ordinary Intentional acts in the same way we look upon the activities of our organs and limbs: we should look upon them as the activity of biological devices whose intended effects define their proper functions. Believing, for instance, is the activity of a device (the brain), which is designed by evolution to have the effect of producing true beliefs in the believer. A true belief is one that is an accurate “map,” or in some sense an accurate account, of how the world is, which in turn enables the believer to find his or her way in the world. Accordingly, Millikan asserts that a given activity produces or gives rise to a content insofar as the activity picks out or refers to something beyond itself. In regard to beliefs, content is revealed, not by the effects of their everyday activities, but by the effects the belief system was designed to have when it is operating in normal conditions. The contents of beliefs are “about” the world in somewhat the same way as accurate maps are “about” the world. Thus, insofar as the contents of our beliefs are dictated by the success of the mapping function they have been designed to perform, “it turns out that we cannot know a priori either that we think or what we think about, just as we do not know a priori whether what we think is true” (1984: 6).

By reference to biology, Millikan believes that she has naturalized Intentionality. Moreover, she points out that this whole approach can be extended to meaning. We can produce a “biosemantics” of sorts, where the meaning of an indicative sentence will be that part of the world, which, in a biological sense, the sentence standardly maps (when true) from the point of view of an interpreter, and the biological function of uttering such sentences will be to map the world accurately (Millikan 1989). But just as the meaning of an indicative sentence is what the sentence properly maps from the point of view of an interpreter, so the Intentionality of a belief is what the belief maps when it is viewed by a believer as the standard effect of a system that is functioning properly in its biologically normal environment.

Thus the Intentionality of a public-language sentence [what the sentence is about] is not derived from the Intentionality of the inner representations [thoughts in the head or other inner signs] that it Normally produces or expresses. Sentences are basic Intentional items ... The position is that [mental-state] Intentionality is grounded in external natural relations, Normal and/or proper relations, between representations and representeds, the notions 'Normal' and 'proper' being defined in terms of evolutionary history—of either the species or the evolving individual or both. Hence nothing that is either merely in consciousness or merely 'in the head' displays Intentionality as such ... Ideas, beliefs, and intentions are not such because of what they do or could do. They are such because of what they are, given the context of their history, supposed to do and of how they are supposed to do it. (1984: 90-3)

As far as Millikan is concerned, our brain states and processes—at least those which give rise to beliefs, desires, etc.—are essentially incomplete when viewed in themselves, since we do not (and cannot) know what they are really “about” until we see the effects of their functioning in the environment and then work out their normal functions and proper effects in an evolutionary sense.

In his book, Mental Content, McGinn extends and defends the teleological-cum-biological account of Intentionality expounded by Millikan. Like Millikan, he takes beliefs and desires to be inner physical states whose Intentional content is a relation between brain structures and things external to these structures. At this point, McGinn’s account meshes closely with Millikan’s, since he agrees that this relation should be explained in terms of proper evolutionary functions. Where McGinn noticeably extends Millikan’s theory is with his contention that “the relational proper
functions of representational mental states coincide with their extrinsically individuated content” (McGinn 1989: 147). In other words, function fixes and is fixed by an internally determinable content. It is because the content of a belief is fixed by reference to its evolutionary proper function that it takes the shape it does—a shape that is apt for “hooking on” the world and meshing with other mental states. The shape in question is a logical one that ultimately dictates the psychological nature of the associated state.

According to McGinn, we would be in error if we look to language for further clarification of the logical shape of Intentional states. Mental states are neither inner conversations nor do they involve inner computations. Logical shape is best explained in terms of the brain’s power to “model,” and the logic in question is the logic built into the brain’s employment of models. As McGinn himself put it:

The basis of content is more like an engineer’s workshop in which no one speaks. There are no volleys of verbal activity occurring in the recesses of your brain, only the production of vastly many practical models. You are not a secret speaker of a hitherto undeciphered language; you are more like the maker of a very sophisticated atlas that covers much more than ordinary geography. (1989:208)

For McGinn, a mechanism is to be described as having content, not merely if it has a normal function shaped by evolution, but also if its function is carried out by generating internal models. This differentiates the kind of content that any bodily organ with an evolutionary proper function might have from the kind of content that brain states manifest.

3.3 Consequences of Externalism

Notice that McGinn’s positing a special power of the brain to produce “practical models” shows some retreat from an externalist position, since the major role of specifying content now seems to have been taken over by an internal item: a model, or representation, or “map” in the head. Such a retreat may turn out to be a blessing, since a purely externalist account of Intentionality has some severe consequences for a number of central notions in the philosophy of mind.

The externalist advances a metaphysical thesis; he rejects the view that individuals who agree in internal physical states must also agree in their psychological explanatory states (states such as belief). That is, he rejects what is now known as ‘psychophysical supervenience.’ The externalist rejection of psychophysical supervenience tells against many popular theories and programs; it tells against all reductionistic accounts of the mental and indeed against token-identity theories. (Owens 1994:138)

Elaborating Owen’s list, we can identify the following consequences—with varying degrees of import—of an externalist psychology.

1. **Type-Identity Must be Abandoned**

Type-identity makes the broad generalization that all mental particulars that could ever exist are identical to neural states. As such, it is a doctrine about folk psychological categories or universals as opposed to specific instances of these categories or universals. Against this view the following functionalist objection has become standard: the psychological constitution of a system seems to depend not on its hardware or physical composition, but on its software or program. So we should not rule out the possibility that a silicon-based organism or a machine can have the same kind
of mental states we do, given that the silicon is properly organized and the machine is correctly programmed. If it is logically possible that silicon-based organisms and machines could develop a psychology comparable to ours, then mental states and neurophysiological processes cannot be identical, however much they prove to be coextensive (Fodor 1981a: 30). But now there is a different reason to reject type-identity. If type-identity is true, then sameness of physiological states should mean sameness of mental states. The thought-experiments discussed above claim to show otherwise, however, that sameness of physiological state does not guarantee sameness of psychological state.

**ii. Token-Identity Must be Abandoned**

Unlike type-identity, token-identity is a theory about the *instances* of folk-psychological categories or universals, not the categories or universals themselves. It states that an agent’s mental state at time *t* is identical to his or her neurophysiological state at time *t*. But according to externalism, this is false. Though Adam and Adam₂ may be physiologically indistinguishable, their thoughts have different contents: one believes he sees water (or has arthritis), the other believes he sees water₂ or has arthritis. The argument, here, is not against the claim that the twins’ mental states are physically realized, but against the identification of a token mental state with a token neurophysiological state; and the reason is a direct consequence of the externalist assumption that more is involved in determining Intentional content than an agent’s biology.

**iii. Weak Supervenience Fails**

In “Thinking Causes,” Davidson defines supervenience as follows: “a predicate *p* is supervenient on a set of predicates *S* if and only if *p* does not distinguish any entities that cannot be distinguished by *S*” (1995: 4). This definition, acknowledged by Davidson to be close to Kim’s (1984) notion of weak supervenience, has sometimes been taken to mean that a difference in mental state is always accompanied by a difference in physical state, but not the converse. The latter formulation implies—in direct contradiction to the externalist position—that the same physical state is always attended by the same mental or psychological state.

**iv. Strong Supervenience Fails**

In “The Myth of Nonreductive Materialism,” Kim attempts to reconcile what we may call the *causal thesis*—i.e., the idea that the mental is causally efficacious and that different psychological states could cause different physical behaviour—with *physical determinism* by appealing to a different kind of supervenience—vis., *strong* supervenience. Kim describes strong supervenience as weak supervenience plus the requirement that there be laws connecting psychological properties with physical properties. But if externalism is correct, it is very difficult to see what sort of laws would permit two distinct mental states to be “attached to” the same physical

---

20 A straightforward version of physical determinism (or the *causal closure of the physical domain*) is: for every physical event *y*, some physical event *x* is causally sufficient for *y*.

21 Kim admits that it can be argued plausibly that supervenience of this strength entails the possibility of reducing the supervenient to the subvenient. However, he stresses that if non-reductive physicalists accept the causal closure of the physical domain, they have no visible way of accounting for psychophysical causation except by giving up their anti-reductionism. Otherwise, they must reject the possibility of psychophysical causal relations.
state. In any case, strong supervenience fails for a much more basic reason: in “Concepts of Supervenience,” Kim points out that strong supervenience entails weak supervenience (but not vice versa). It follows that since psychological externalism means the demise of weak supervenience, it must also mean the demise of strong supervenience.

**Global Supervenience**

It is important to note, however, that though weak and strong supervenience are incompatible with externalism, *global* supervenience may not be. Global supervenience states that for any two worlds, *A*-respects (e.g., the mental) globally supervene on *B*-respects (e.g., the physical) if and only if the two worlds’ being *B*-twins (i.e., internally physically identical) means that they are also *A*-twins (i.e., have exactly the same pattern of distribution of mental properties). This type of supervenience seems consistent with all the externalist arguments discussed above. Putnam’s thought-experiment, for example, alleges that Earth and Twin-Earth are not *A*-twins (have a different pattern of distribution of mental properties) precisely because they are not *B*-twins (because they differ physically in respect to what their respective citizens call “water”). But if the two planets were *B*-respects (physically) identical, Adam and his doppelganger would not differ in their propositional attitudes.

Global supervenience also seems compatible with the kind of externalism endorsed by Davidson and Millikan. Assuming we can consider the causal history of the events of a given world as that world’s *B*-respects, we can easily adapt global supervenience to aetiological-motivated anti-individualism. In this case, the relevant externalist claim would be: a difference in two mental states, *m* and *n*, implies a difference in the causal histories of *m* and *n*. We can now agree with Davidson and Millikan while at the same time upholding a version of global supervenience that substitutes causal history for physical constitution. In other words, we can agree that the content of Intentional states is fixed by the aetiology of those states and allow that for any two worlds, mental states (*A*-respects) globally supervene on the causal history of those states (*B*-respects) if and only if the two worlds’ being *B*-twins (i.e., identical in the aetiology of their events) means that they are also *A*-twins (i.e., have exactly the same pattern of distribution of mental properties).

A similar substitution makes global supervenience amenable to Burge’s variety of externalism. Assuming that we can consider the sociolinguistic habits of a world as that world’s *B*-respects, the claim now would be: a difference in the mental states of two individuals, *A* and *B*, implies a difference in the relevant sociolinguistic practices of the communities of which *A* and *B* are members. So we can consistently hold that the Intentionality of a person’s mental state is decided by the sociolinguistic habits of that person’s community and that for any two worlds, mental states (*A*-respects) globally supervene on sociolinguistic practices (*B*-respects) if and only if the two

---

22 For purposes of argument, “causal history” may be taken to indicate any time-frame (including the span of evolution).

23 Dretske (1995: 161-2) seems to suggest just such a view: “Whether or not the mental has to supervene on the physical constitution of a system in order to be causally relevant to that system’s behavior depends ... on what one takes to be the behavior of the system. If behavior is identified, not with bodily movements and change, but with the causal processes that result in bodily movement and change, then ... one would expect the content of thought and the quality of experience to supervene not on the physical constitution of the animal but on the historical events and processes that shaped the animal’s current control circuits ... ” This, of course, is no argument for global supervenience. Such an argument would require an explanation of why the identification of behavior with the *history* of causal processes that result in bodily movement or change at time *t* is more viable than adopting a more temporally localized (e.g., token-based) psychophysical supervenience, given that at time *t* the animal’s control circuitry has already been shaped by evolution and is in place.

24 I think the word “relevant” is important here, because it stands to reason that not all sociolinguistic practices must play a role in the determination of Intentional content.
v. Mind-Brain Reductionist Programs are Out of Reach

If externalism is correct, mind-body identity is not permitted and neither is psychophysical supervenience. So unless there be some way of figuring out how to reduce the mind to the brain without making use of these strategies, it appears that any mind-body reductionist ambition will be frustrated. This conclusion holds regardless of what we think reduction is and what conditions we think are necessary for it to occur. The reason is that, if the externalists get their way, Intentional content will always involve more than the physical states of brains; it will involve one or another agent-environment relation. So the brain, to put it bluntly, will not be enough!

vi. Mental Causation is Jeopardized

Adherents to mind-body causation normally reconcile the causal thesis with the causal closure of the physical domain by adopting some form of psychophysical identity or supervenience. Now that these routes have been closed, the causal efficacy of the mental appears to have lost its footing. Hence Fodor warns: “Causal powers supervene on local microstructure. In the psychological case, they supervene on local neural structure. We abandon this principle at our peril; mind/brain supervenience (/identity) is our only plausible account of how mental states should have the causal powers that they do have” (1987a: 44).

The trouble for mental causation is best revealed by Davidson’s Swampman. By hypothesis, Swampman is as behaviourally similar to Davidson as you like: not only is he physically identical to Davidson, he walks and talks like him; he seems to have the same tastes in fashion, food, music, and literature; he seems to hold the same sociopolitical views, the same philosophical opinions and doctrines; he seems to react in the same way to Davidson’s family, friends, and colleagues, so that “no one can tell the difference.” Yet Swampman has no mental states at all. If this be the case, we might wonder what possible difference do mental states make. Swampman, if indeed he is a zombie, seems to do just as well as Davidson without them. As Heil (1994: 160) notes, “were my history and circumstances such that the states responsible for my deeds exhibited very different Intentional contents, or even no contents at all, they would nevertheless produce identical bodily movements. Or at least it is hard to imagine why they should not.”

vii. No Privileged Access to First-Person Mental States

There is a current controversy as to whether or not direct knowledge of personal Intentional states is inconsistent with anti-individualism. I think it is; others don’t. Interestingly, Burge and Davidson are among the compatibilists.²⁶

²⁵ Neither Burge nor any other social externalist must hold this position, but they certainly can do so while remaining faithful to their anti-individualism.

The problem is often phrased in terms of an *asymmetry of epistemic access* between first-person and third-person mentality. The claim is that in the first-person case, no empirical evidence is required for knowing that one is having such-and-such thought, whereas this kind of evidence is always necessary for achieving a similar knowledge in the case of others. The difference turns on the *manner* in which the respective thoughts are arrived at or justified. I have no choice but to *infer* what those around me think from observations about what they do and say. In my own case, however, inference is neither necessary nor relevant. I know what I think without appeal to supplementary evidence. Even when such evidence is available, I seldom ever consult it.

Putting the issue in this way does not commit us to either *infallibility* or *incorrigibility* with respect to the mental. Infallibility requires that whenever we exemplify a given mental property, we know that we do so; incorrigibility requires that the beliefs we harbour concerning our own states of mind cannot fail to be true. We have reason, I think, to reject both doctrines. Our readiness to accept the ideas of repression and the unconscious, together with our willingness to acknowledge that we can fail to realize our deepest preferences and prejudices, suggest that incorrigibility and infallibility are notions better left behind. Fortunately, neither position is entailed by the claim that there is an asymmetry of epistemic access between first-person and third-person psychology. One can acknowledge a difference in epistemic access while still holding that we sometimes make mistakes about what we think, just as we sometimes make mistakes about the thoughts of others.

Burge agrees that some mental states, particularly those falling under Descartes’ paradigm—e.g., “I am now thinking,” “I doubt that this very thought can be an illusion,” “I judge (or doubt) that water is more common than gold,” etc.—constitute real self-knowledge, “that they are not products of ordinary empirical investigation, and that they are peculiarly direct and authoritative” (1988: 112). Davidson expresses a similar attitude: “While it is clear enough, at least in outline, what we have to go on in trying to fathom the thoughts of others, it is obscure why, in our own case, we can so often know what we think without appeal to evidence or recourse to observation ... Thus sincere first-person present-tense claims about thoughts, while neither infallible nor incorrigible, have an authority no second or third person claim ... can have” (1987: 88).

The problem for Burge and Davidson seems to be one of understanding how we can know some of our mental events directly, in a manner that does not involve “ordinary empirical investigation,” when those events depend for their identities on our relations to the environment. In other words, why is our having non-empirical knowledge of our thoughts not impugned by the fact that such thoughts are individuated by relations to the environment that we can know only empirically?

Davidson suggests that the alleged difficulty partly stems from the incorrect assumption that, if a thought is identified by a relation to something outside the head, it is not wholly in the head. He argues that this is no more true than the claim that because a sunburn presupposes the existence of the sun, a person’s sunburn is therefore *not* a condition of his or her skin. “Individual states and events,” says Davidson, “don’t conceptually presuppose anything in themselves; some of their descriptions may, however” (1987: 103). He alleges that philosophers who find difficulty in situating privileged access within externalism are confusing *thoughts* with thought-*descriptions*: they make the mistake of inferring from the fact that a thought is identified or described by relating it to things outside the head, that the thought itself must be outside the head, thus unavailable to privileged access. Davidson denies this; he permits that a person may well have authoritative knowledge of his or her mental states, though not necessary under all possible descriptions of those states.
Davidson apparently wishes to claim that one could have privileged access to an episode of thought independently of having privileged access to any particular description the episode might satisfy. But assuming that descriptions are not those things to which we have privileged access, just what is it that we have privileged access to? Subtracting the descriptions of a mental episode from the episode itself leaves the privilege of knowing only that the episode exists. But surely having this sort of access is not much of a privilege. The traditional view is not just that we have direct access to the fact that our thoughts occur; rather, the view is that we have direct access to our thoughts as satisfying certain descriptions. In particular, the traditional view is that we have privileged access to our thoughts as having certain contents. So Davidson's defence of privileged access is misleading at best (McKinsey 1991: 177).

In contrast to Davidson, Burge endeavours to defend privileged access in its traditional guise. Assume that Oscar is an English-speaking Earthian who does not himself know the chemical constitution of water, but is part of a community, some of whose members do know that water is composed of H₂O. Burge's arguments in "Individualism and Self-Knowledge" and "Other Bodies" commit him to the claim that the following three propositions are consistent:

1. When Oscar thinks that water is wet, he knows a priori that he is thinking water is wet.²⁷
2. The proposition that Oscar is thinking that water is wet necessarily depends upon ξ.  
3. ξ cannot be known a priori, but only by empirical investigation,²⁸

where ξ stands for the proposition asserting the existence of the entities which are entailed by Oscar's having the thought that water is wet. The origin of (1) is what Burge (1988: 112) calls "basic self-knowledge"—vis., that subjects know some of their thought contents in a direct, non-empirical manner; the basis of (2) is Burge's claim that "... an individual's having certain de dicto attitudes entails the existence of entities other than himself and his attitude contents" (1982: 117);²⁹ and the basis of (3) is the claim that a subject can only have empirical knowledge of the environmental facts implicated in the individuation of his or her Intentional states (1988: 112).

Notice that if (2) is knowable a priori, then (1)-to-(3) are inconsistent: since Oscar knows a priori that he is thinking water is wet, if he also knows (2) a priori, then he is capable of knowing ξ a priori; but by (3), ξ is not knowable a priori. It seems, then, that the question of the consistency of (1)-to-(3) depends on whether or not a subject can know (2) a priori. In "Anti-Individualism and Privileged Access," McKinsey argues that one can indeed have a priori knowledge of (2), and therefore, that externalism and privileged access are incompatible.

McKinsey argues that (2) is knowable a priori by claiming that in (2), the notion of "necessary dependence" must be interpreted as conceptual implication, not metaphysical necessity. Thus externalism must be the thesis that the possession of certain mental properties (e.g., the property of thinking that water is wet) conceptually (not metaphysically) implies the existence of objects external to the person who has the mental property. Following this interpretation, Burge's thesis of anti-individualism should be understood as:

²⁷ In the present context, Burge takes a priori to mean, roughly, "without the need for empirical evidence."


²⁹ And on the previous page of the same article, Burge makes the following claim about Adam "... all of Adam's attitude contents involving relevant natural kind notions—and thus all his relevant attitudes (whether de re or de dicto)—are individuated, by reference to other entities."
Some cognitive states that are described by *de dicto* attitude sentences (e.g., “Oscar is thinking that water is wet”) conceptually imply the existence of objects external to the person to whom the state is ascribed.

and (2) must be understood as:

\[(2b)\] The proposition “Oscar is thinking that water is wet” conceptually implies \(\xi\).

It is now easy to see that (1), (2b), and (3) make an inconsistent triad. The argument is this: by (1) Oscar knows *a priori* that he is thinking that water is wet; by (2b) Oscar can simply *deduce* \(\xi\) using only premises that are knowable *a priori* (including the premise that he is thinking that water is wet), because the relationship between (1) and (2b) is now one of *conceptual* entailment not requiring empirical investigation. Since Oscar can deduce \(\xi\) from premises that are knowable *a priori*, Oscar can know \(\xi\) itself *a priori*. But this contradicts (3). So (1), (2b), and (3) are inconsistent.

Adopting a metaphysical interpretation of “necessary dependence” would, in McKinsey’s view, make (1), (2), and (3) consistent: since metaphysical dependencies are often only knowable *a posteriori*, propositions that are knowable *a priori* might metaphysically depend upon other propositions that are only knowable *a posteriori*. But viewing necessary dependence in this way trivialises anti-individualism; for anti-individualism is the thesis that some *de dicto* cognitive-attitude states are wide, and to say that a state is wide cannot merely mean that the state metaphysically entails the existence of external objects. If it did, then given certain materialist assumptions commonly held,\(^{30}\) it would follow that all psychological states of *any* kind would be wide, so that externalism would merely be a trivial consequence of token-physicalism.

But this is obviously *not* the sense of “wide psychological state” that Putnam, Davidson, or even Burge himself had in mind. While it may be true that Oscar’s thinking that water is wet entails the existence of Oscar’s mother, or the existence of the egg and sperm from which Oscar originated, it is not for *this* kind of reason that Oscar’s mental state is wide. These considerations have led McKinsey to conclude that externalism should be stated in terms of conceptual implication rather than metaphysical implication; and this, it turns out, means that externalism is inconsistent with first-person authority.

I have concentrated mainly on the compatibilist arguments of Burge and Davidson, but the debate is much wider in scope than I am able to review in this section.\(^{31}\) As indicated, my own attitude is that privileged access is incompatible with externalism; I have not yet come across a convincing argument to the contrary. I shall revisit the issue again in chapter VII to see what sense our scriptal approach to Intentionality can make of it.

### 3.4 Three Individualist Offerings

So far we have focused almost exclusively on externalism. I want now to turn my attention to the internalist perspective, concentrating on the theories of (early) Jerry Fodor, John Searle, and Gabriel Segal. In so doing, I hope to show how some have defined Intentionality as a property of

---

30 For instance, that no human could (metaphysically) have existed without biological patterns donated or replicated, in one way or another, by other human beings.

31 For an overview, see Ludlow and Martin's *Externalism and Self-Knowledge*. 
mental states located entirely “in the head,” thereby providing a contrast with the viewpoints discussed above.

i. Fodor

Dennett argued that when we are said to believe that it is now raining, our heads do not contain states or processes that encapsulate, or encode, or represent in any way the sentence or proposition, “It is now raining;” we merely find it useful to speak as if they do. The “language of thought” hypothesis, on the contrary, instructs that when we believe that it is now raining, our heads do contain, in a perfectly literal sense, the proposition, “It is now raining.” In the mid 1970s, Jerry Fodor set out to defend this position.

If Fodor can be said to have had an intellectual mentor, it is Noam Chomsky. Chomsky believed that all natural languages make use of more or less the same formal operations in generating well-formed or grammatical sentences. In this sense, all human languages are remarkably similar in structure, and so share a “universal grammar,” or are circumscribed by the same “general linguistic theory”:

The study of universal grammar ... is the study of the nature of human intellectual capacities. It tries to formulate the necessary and sufficient conditions that a system must meet to qualify as a potential human language, conditions that are not accidentally true of the existing human languages, but are rather rooted in the human language capacity, and thus constitute the innate organization that determines what counts as linguistic experience and what knowledge of language arises on the basis of this experience. (Chomsky 1968: 24)

Chomsky suggests that the most plausible explanation of this universal grammar is that human children are genetically endowed with an innate ability to engage in those formal operations that enable them to learn whatever natural language is spoken in their native environment. He believes that positing such a universal grammar makes sense of the fact that children learn to speak in a reasonably grammatical way even after a brief exposure to spoken samples of their native tongue. For Chomsky, this constitutes evidence that the universal grammar is “wired in” the human brain (1968: 58-61).

At one point, Fodor says that “Chomsky’s demonstration that there is serious evidence for the innateness of what he calls ‘General Linguistic Theory’ is the existence proof for the possibility of a cognitive science” (1981b: 258). As a functionalist, Fodor holds that to describe humans as having beliefs, desires, hopes, etc., is to describe the cognitive functions of human beings. Unlike Dennett’s instrumentalist functionalism, however, Fodor takes these descriptions literally: for him, people really do have beliefs, and desires, and hopes in their heads. In other words, Fodor is an industrial-strength realist about Intentionality and the propositional attitudes. He is adamant that any psychological explanation of the sources of human behaviour must ultimately make reference to, or at least presuppose, a human’s employment of an internal representational system of very considerable richness. Since—unlike Brentano and latter-day Brentanians—Fodor does not believe that we have grounds for positing any mental substance independent of our physical organs, he argues that this representational system must be not merely the language of thought, but the language of the brain.

The crux of Fodor’s theory is that Intentionality is primarily and originally a feature of brain states. Language is Intentional only in a secondary sense, only insofar as some of the sentences we
utter describe real features of our minds and the minds of others.\textsuperscript{32} It is this fact that leads us to generate mentalistic vocabulary. Accordingly, if I say that Delilah decided such-and-such because she believed certain things, had certain desires, connected the two in her mind and came up with a certain evaluative decision, I am describing a series of real processes in Delilah's brain—processes that involve complex computational operations over encoded propositional contents.\textsuperscript{33} Fodor argues that because computation can take place only after information has been encoded in computable form, there can be no computation without representation (1975: 31).

Fodor and Chomsky depart in one important respect. Chomsky argued that we have strong grounds, from empirical data about natural languages and how we come to learn them, for positing a universal grammar or universal capacity for generating our different, yet importantly similar, natural dialects. Fodor, on the other hand, believes that we have strong empirical grounds for positing a real innate language or representational system which is the basis for our universal capacity to learn natural languages, acquire concepts, and even perceive in the full sense. In short, Fodor is arguing for a representational theory of mind (RTM), which makes empirical claims about how our brains operate and so runs the risk of being falsified by future research.\textsuperscript{34}

With respect to Intentionality, RTM entails that any propositional attitude is literally a computational relation between an organism and some formula in the internal code of that organism. So "to believe that such and such is to have a mental symbol that means that such and such tokened in your head in a certain way" (Fodor 1987a: 17). Moreover, it is in virtue of this system of representing and processing information that mental states are related causally to one another. In Psychosemantics, Fodor develops his account of how the brain, in pursuing its purely mental activities, operates causally and Intentionally simultaneously, and in such a way that the causal paths of brain processes are the paths directed by the rational interplay of contents represented by those processes.

To a first approximation, to think 'It's going to rain; so I'll go indoors' is to have a tokening of a mental representation that means I'll go indoors caused, in a certain way, by a tokening of a mental representation that means It's going to rain ... The trick is to combine the postulation of mental representations with the 'computer metaphor.' Computers show us how to connect semantical with causal properties for symbols. So, if having a propositional attitude involves tokening a symbol, then we can get some leverage on connecting semantical

\textsuperscript{32} Searle (1983) also accepts the primacy of the mental, as does Stalnaker, who remarks: "I have argued, and continue to believe, that what I have called 'the linguistic picture'—the family of doctrines, metaphors and strategies that assume, in one way or another, the priority of linguistic over mental representation—has had a profound influence on our conception of Intentionality, a distorting influence that has impeded a clear understanding not only of thought, but also of speech" (1999: 1-2). On the other side we find Sellars (1956) and Dummett (1991), both of whom think that the Intentionality of the mental derives from the Intentionality of language. Dummett claims that "the philosophy of thought can be approached only through the philosophy of language. That is to say, there can be no account of what thought is, independently of its means of expression" (1991: 3-4). Burge too, I think, must be included in the second group; see particularly his (1977). Interestingly, Dennett rejects both the "primacy of mind" and the "primacy of language" hypotheses, claiming instead that it is "Mother Nature," or evolution by natural selection, that has intrinsic Intentionality, and that naturally intelligent psychological beings—humans and other complex subjects—exhibit only derivative Intentionality. In "Evolution, Error, and Intentionality," Dennett concludes: "If there is to be any original Intentionality—original in the sense of being derived from no other, ulterior source—the Intentionality of natural selection deserves the honor" (1987a: 318).

\textsuperscript{33} Fodor defines "a computation" as "a transformation of representations which respects ... [the representations'] semantic relations" (1983: 5).

\textsuperscript{34} In The Modularity of Mind, Fodor develops RTM in the direction of suggesting that the mind-cum-brain is not merely representational and computational, but also functions by means of a series of interconnected "modular systems" or "special purpose computational mechanisms" (1983: 120). Psychological evidence certainly supports this conception, at least as far as memory is concerned; see below, §5.2.i.
properties with causal ones for *thoughts* ... Here, in barest outline, is how the new story is supposed to go: You connect the causal properties of a symbol with its semantic properties *via its syntax*. The syntax of a symbol is one of its higher-order physical properties ... [W]e can think of the syntactic structure of a symbol as an abstract feature of its shape ... It's easy, that is to say, to imagine symbol tokens interacting causally *in virtue of* their syntactic structures. The syntax of a symbol might determine the causes and effects of its tokenings in such a way that the geometry of a key determines which locks it will open. (1987a: 17-19)

Fodor points out that an important advantage of the Representational Theory of Mind is the preservation of the autonomy of psychology. If RTM gives us an accurate picture of how the brain works, then the proper explanations in psychology will be computational ones in terms of the interplay of the propositional attitudes, which, in turn, will be explained in terms of the various ways of processing contents as represented in the language of thought. At both these levels the psychological explanation is produced in Intentional terms. It would follow, then, that psychological explanations could never be reduced to the non-Intentional vocabulary of neurophysiology:

Quite possibly there will never be a state of science [when] we can, as it were, do neurology instead of psychology because, quite possibly, it will never be possible to express in the vocabulary of neurology those generalizations about relations of content that computational psychological theories articulate. Psychologists have lots of things to worry about, but technical unemployment is not likely to be one of them. (1981b: 165)

Fodor believes that RTM also bestows a sort of autonomy on mental states *themselves*, in the sense that these states can (and should) be investigated independently of, and so without reference to, the context which may have given them rise, and the behavioural output which they in turn might produce. In other words, what Fodor has in mind as the correct approach to psychological research is what he calls *methodological solipsism*, which focuses on mental states and processes in single-minded isolation, and entails that these states and processes are wholly individuated by exclusive reference to items internal to the organism whose states or processes they are.

To put the matter in another way, suppose that the semantic evaluation of a mental state depends on certain of its relational properties—in effect, on how the state corresponds to the world. If this is the case, then methodological solipsism is simply the doctrine that psychological states are individuated *without respect to their semantic evaluation*. This intuition is also driving Stephen Stich's *principle of psychological autonomy*: "the properties and relations to be invoked in an explanatory psychological theory must be supervenient upon the current, internal physical properties and relations of organisms (i.e., just those properties that an organism shares with all its replicas)" (Stich 1978: 347). For Fodor, there is neither need nor justification for any investigation into the environmental causes or behavioural effects of the states or processes in question. To think otherwise is to fall back into behaviourism.

---

35 Recall that Quine's reaction to this Fodor-Brentano view was a wholesale rejection of Intentional idioms.

36 See Fodor (1980). In Fodor (1987a: ch. 2) methodological solipsism is contrasted with *methodological individualism*, which asserts that psychological states are individuated *with respect to their causal powers*. The difference between solipsism and individualism is that the latter does not prohibit the relational individuation of mental states; it just says that no property of mental states, relational or otherwise, counts taxonomically unless it affects causal powers. Fodor credits the term "methodological solipsism" to Hilary Putnam (1975).

37 Unlike Fodor, however, Stich uses the principle of psychological autonomy to attack folk psychology. See Yagisawa (1985) for a criticism of Stich's use of this principle.

38 It is actually rather hard to get a firm hold on just what Fodor's position is, probably because his views have undergone some rather drastic shifts. So, for instance, near the end of *Psychosemantics* we encounter what seems to be a sort of epiphenomenalism: "I don't believe that there are Intentional mechanisms. That is, I don't believe that contents per se determine causal roles. In consequence, it's got to be possible to tell the whole story about mental causation (the whole
ii. Searle

Searle’s theory of Intentionality is explicated with the aid of an alleged analogy between Intentional states and speech acts. Because Searle, like Fodor, believes that Intentionality is originally a feature of the mind, he is careful to observe that his explaining Intentionality in terms of speech acts does not imply that Intentional states are essentially linguistic. In fact, he argues for the opposite claim: “Since sentences—the sounds that come out of one’s mouth or the marks that one makes on paper—are, considered in one way, just objects in the world like any other objects, their capacity to represent is not intrinsic but is derived from the Intentionality of the mind. The Intentionality of mental states, on the other hand, is not derived from some more prior forms of Intentionality but is intrinsic to the states themselves” (1983: vii). So while the direction of Searle’s pedagogy is to explain Intentionality in terms of language, the direction of his theorizing is to explain language in terms of Intentionality.

Searle maintains that there are at least four points of similarity or connection between Intentional states and speech acts (1979: 75-80). First, he argues that the difference between propositional content and illocutionary force—a difference familiar within the theory of speech acts—carries over to Intentional states. Speech-acts theorists commonly distinguish between the propositional content that, for instance, Eva will leave the room, and the illocutionary force with which that propositional content is expressed. In the case of Intentional states, there is a parallel distinction between the representative content that Eva will leave the room and the psychological mode—i.e., believing, hoping, desiring, or fearing—in which one has that representative content.

Searle next claims that the idea of directions of fit, also familiar in speech theory, is manifest in Intentional states. Different classes of speech acts serve to show how this notion operates. The members of the assertive class of speech acts—statements, descriptions, assertions, etc.—are supposed to match an independently existing world. To the extent that they do so or fail to do so, they are said to be true or false. Hence the assertive class of speech acts has a word-to-world direction of fit. On the other hand, members of the directive class of speech acts—orders, commands, requests—and members of the commissive class—promises, vows, pledges—are not supposed to match an independently existing reality, but are supposed to bring about changes in the world so the world matches the speech act. To the extent that they succeed or fail to do so, we do not say that they are true or false, but that they are obeyed or disobeyed, fulfilled or unfulfilled, kept or broken. Accordingly, the directive and commissive classes of speech acts have a world-to-word direction of fit. Searle argues that something very much like these categories are discernible in Intentional states:

If my beliefs turn out to be wrong, it is my beliefs and not the world which is at fault, as is shown by the fact that I can correct the situation simply by changing my beliefs. It is the responsibility of the belief, so to speak, to match the world, and where the match fails I repair the situation by changing the belief. But if I fail to carry out my intentions or if my desires are unfulfilled I cannot in that way correct the situation by simply changing the intention or desire ... Beliefs like statements can be true or false, and we might say they have the 'mind-to-world' direction of fit, whereas desires and intentions cannot be true or false, but can be complied with or unfulfilled. We might say that they have the 'world-to-mind' direction of fit. (1979: 77)

story about the implementation of the generalizations that belief/desire psychologies articulate) without referring to the Intentional properties of the mental states that such generalizations subsume” (1987a: 139). Later, in The Elm and the Expert, Fodor expressly accepts an externalist semantics and attempts to reconcile his new position with the language of thought hypothesis by eliminating narrow content altogether. He defines a thought as a three-place relation between a thinker, a broad content, and a mode of presentation (i.e., a sentence of Mentalese), “and since linguistic expressions are individuated (inter alia) by their syntax, token thoughts are type distinct if they differ either in their [broad] contents or in their modes of presentation” (1994: 55).
Searle finds a third connection between Intentional states and speech acts in the fact that in the performance of any illocutionary token with a propositional attitude, we express a certain *Intentional state* with a corresponding propositional content, and the Intentional state thus expressed is the *sincerity condition* for that type of speech act. Thus if one makes the statement that $m$, one expresses a belief that $m$; if one promises to do $p$, one expresses an intention to do $p$; if one orders someone to do $y$, one expresses a desire that (s)he should do $y$, and so on. Moreover, the connection between illocutionary acts and the Intentional sincerity conditions of these acts is, on Searle’s view, *internal*. That is to say that “the performance of the speech act is necessarily an expression of the corresponding Intentional state, as is shown by Moore’s paradox. You can’t say, ‘It’s snowing but I don’t believe it’s snowing,’ ‘I order you to stop smoking but I don’t want you to stop smoking,’ ‘I apologize for insulting you, but I am not sorry that I insulted you’ …” (1979: 78).

Finally, Searle argues that the notion of *conditions of satisfaction* applies quite generally across speech acts and Intentional states. We say that a statement is true or false, that an order is obeyed or disobeyed, that a promise is kept or broken. In each of these cases we ascribe success or failure based on the direction of fit between word and world. But this notion of satisfaction clearly applies to Intentional states as well. My belief is satisfied when things are as I believe them to be; my desire is satisfied when it is fulfilled; my intention is satisfied when it is carried out. Searle points out that in general the speech act will be satisfied if and only if the expressed psychological state is also satisfied. Thus my statement will be true if and only if the expressed belief is correct; my order will be obeyed if and only if the expressed wish or desire is accomplished; my promise will be kept if and only if the expressed intention is carried out. Ultimately, then, the conditions of satisfaction of the speech act and the conditions of satisfaction of the expressed Intentional state are identical.

Searle believes that his approach to Intentionality enables us to see the way to a solution for several outstanding problems. One of these problems concerns the question “What is an Intentional state?” which, according to him, need not be construed ontologically. He claims that what makes a mental state Intentional is not its ontological category, but its logical properties, so that the traditional ontological problems about mental states are simply irrelevant to their Intentional features. The ontological problems relating to the status of Intentional *objects* also receive a simple solution: “An Intentional object is just an object like any other; it has no peculiar ontological status. To call something an Intentional object is just to say it is what some Intentional state is about” (1979: 82). If this is true then it would be a mistake to say that an Intentional state—e.g., a belief—is a two-term relation between a believer and a proposition. An analogous mistake would be to say that a statement is a two-term relation between a speaker and a proposition. One should rather say that a proposition is, not the *object* of a statement or belief, but its *content*. So the content of the belief that Delilah is Samson’s pen-pal is the proposition that Delilah is Samson’s pen-pal, but that proposition is not what the statement or belief is about.

Another alleged advantage of Searle’s theory is that it provides a simple account of the relationship between Intentionality (*with-a-t*) and intensionality (*with-an-s*):

One of the most pervasive confusions in contemporary philosophy is the mistaken belief that there is some close connection, perhaps even an identity, between Intensionality-*with-an-s* and Intentionality-*with-a-t*. Nothing could be further from the truth. They are not even remotely similar. Intentionality-*with-a-t* is that property of the mind by which it is able to represent other things; Intensionality-*with-an-s* is the failure of certain sentences, statements, etc. to satisfy certain logical tests of extensionality. The only connection between them is that some sentences about Intentionality-*with-a-t* are Intensional-*with-an-s*. (1979: 85)

Now, traditionally the puzzle about intensional sentences is how it can be the case that their use to make statements does not permit the standard logical operations if (as seems to be the case) the
words contained in the sentences carry their normal meaning, and if the logical properties of a sentence are a function of its meaning and the meaning of a sentence is in turn a function of the semantic properties of its component parts.

The answer suggested by Searle’s approach is that since the sentence “Samson believes that his pen-pal is coming to visit” is used to make a statement about an Intentional state—namely, Samson’s belief—and since an Intentional state is a representation, then the sentence is a representation of a representation. As such, the truth conditions of the sentence will depend on the features of the representation being represented—in this case, the features of Samson’s belief—and not on the features of the objects or states of affairs represented by Samson’s belief (including the identity of his visitor). We have, therefore, no guarantee that the substitution of key terms with their extensional equivalents will leave unaffected the truth-value of the sentence as a whole: it may simultaneously be true that Samson believes his pen-pal will visit and false that he believes Delilah will visit.

Notwithstanding all the similarities, there is an obvious disanalogy between mental states and speech acts: mental states are states, speech acts are acts. According to Searle, this difference has an important consequence for the way language is related to its physical realization. The actual performance in which a speech act is made involves the production or use of some physical entities, such as vocalized noises or marks on paper. Beliefs, hopes, desires, and fears, on the other hand, are intrinsically Intentional. To call certain states “beliefs,” “hopes,” “desires,” or “fears” is already to ascribe Intentionality to them. But speech acts have a physical level of realization, qua speech acts, that is not Intrinsically Intentional: there is nothing inherently Intentional about the noises that come out of one’s mouth or the marks one sets on paper. The question now is: how does the mind impose Intentionality on entities that are not intrinsically Intentional, on entities like sounds or marks “that are, considered in one way, just objects in the world like any other objects”?

Searle’s answer is that there is a double level of Intentionality in the performance of speech acts. There is, first, the Intentional state expressed; secondly, there is the intention (in the ordinary sense of the word) with which the utterance is made. Searle claims that it is this second Intentional state that bestows Intentionality on physical phenomena. That is, the mind imposes Intentionality on objects that are not inherently intentional by transferring the conditions of satisfaction of the expressed psychological state to the external physical entity.

The double level of Intentionality in the speech act can be described by saying that by intentionally uttering something with a certain set of conditions of success, those that are specified by the essential condition for that speech act, I have made the utterance Intentional, and thus necessarily expressed the corresponding psychological state. I couldn’t make a statement without expressing a belief or make a promise without expressing an intention because the essential condition on the speech act has as conditions of satisfaction the same conditions of satisfaction as the expressed Intentional state. (1979:89)

For Searle, the four connections between mental states and speech acts suggest a certain picture of Intentional states as consisting of a representative content in a certain psychological mode; these are the two essential ingredients of Intentionality. Moreover, Searle takes Intentional states to represent objects and states of affairs in exactly the same way that speech acts represent objects and states of affairs; the only difference being that the mind, unlike language, has its Intentionality essentially. This means that we need not resort to any social, or otherwise external, facts in fixing or individuating Intentional states: both the representative content and psychological mode are “inner,” or “in the head,” and do not, as such, necessarily presuppose the existence of anything outside the agent who does the representing.
None of this is meant to be a true analysis of Intentionality, however—at least not in the classic philosophical sense of giving necessary and sufficient conditions in terms of simpler notions—and Searle readily acknowledges this (1983: 79). If his characterization of Intentionality in terms of representation were intended to be a genuine analysis in the classic sense, then it would be hopelessly circular; for the idea of representation is as much Intentional as that which it was intended to illuminate. But Searle believes that it is impossible to give an analysis of Intentionality as such, for “any attempt to characterize Intentionality must inevitably use Intentional notions, and thus any such attempt will move within ... the circle of Intentionality” (1979: 90).

iii. Segal

A rather strong version of individualism has been recently defended by Gabriel Segal in *A Slim Book About Narrow Content*. Segal starts by observing that a good understanding of a property requires knowing whether that property is relational or intrinsic. What makes a property relational is that an object’s possession of it depends not only on the object itself, but also on circumstances external to the object. Obvious examples include residing in Alexandria, weighing more than Oprah Winfrey, being a parent or a nephew. Intrinsic properties, in contrast, include such things as chemical and physical constitution, both of which do not depend on the relation of the object to its environment. Segal’s concern is whether the *cognitive content* of certain psychological states is constituted relationally or intrinsically. By “cognitive content” he means those properties that account for the role of mentality in typical psychological predictions and explanations.

Segal believes that content supervenes on local microstructure—that is, if two beings are identical with respect to their physiology, they *must* also be identical with respect to their cognition.39 His main thesis is that being in a contentful psychological state does not essentially involve standing in any relation to external objects or conditions. He attempts to make his case in three steps—criticizing, first, the two leading brands of externalist theories: those involving natural kinds and those based on social dynamics. Next, he rejects popular two-factor views, which posit a narrow as well as a broad content. Finally, he offers a radical version of internalism—arguing that narrow content, a variety of ordinary representation, is all there is to Intentionality. In thus defending internalism, Segal does not claim to espouse a general philosophical theory. At this stage, he suggests that it should suffice simply to shed reasonable doubt on anti-individualism and provide reasons to believe that good psychology is, or could be, internalistic.40

**Natural Kinds and Social Externalism** With respect to natural-kind concepts, one externalist position that has been motivated by Putnamian Twin-World considerations is what Segal (2000: 30) calls “the thesis of world dependence of kind terms” (TWD). TWD has three characteristics:

First, it means that the extension conditions of a nonempty kind concept depend in part on a real relationship between thinkers and samples in the external world. So thinkers must have interacted with samples or know someone who has interacted with samples or been in some other form of direct or indirect causal contact with the

---

39 The “must” is interpreted nomologically, not metaphysically. “Thus local supervenience is the thesis that microstructure nomologically determines cognitive properties, that twins are nomologically possible twins, and so on” (2000: 17).

40 These kinds of qualifications have become standard in the philosophy of mind. No one claims to have definitive answers anymore, or to provide knockdown arguments. Whether this is indicative of our poor state of knowledge or symptomatic of a general prudence due to increased broadmindedness, is not an easy question to answer. I am inclined to think that both factors are at work, with the former being somewhat more prominent than the latter.
samples. Second, it means that extension conditions are essential to the cognitive content of the concept. So if concept C is world-dependent, then any concept that has different extension conditions from C has a different cognitive content from C. The third characteristic of world dependence rules out decomposition of content into two factors. (2000:30-1)

According to Segal, empty concepts provide the largest problem for TWD. One way to get at the difficulty is to consider the fate of a kind term in the counterfactual circumstance in which the term has no extension. Suppose there were no water—or aluminium, or topaz, or quarks, or polio—but that otherwise things are as much as they are as could be. What would happen to the words “water,” “aluminium,” “topaz,” “quarks,” and “polio”? What concepts would these terms express? The thesis of world dependence entails that either they express no concepts at all, or they express different concepts from the ones they actually express. But neither alternative, claims Segal, is acceptable to the externalist.

Segal’s argument against the first option is straightforward: there do exist plenty of empty kind terms—terms that express concepts yet lack extension. In fact, far from being a rarity, empty concepts are a pervasive and significant feature of human communication. As examples, Segal cites the concepts of god, spirits, ghosts, ether, and phlogiston. He contends that the point of attributing conceptual content in all these cases is simply that only by so doing can we make psychological sense of a wide variety of human activity. Anthropologists, historians of culture and science, psychoanalysts, and others do so readily, thereby constructing what appear to be perfectly cogent psychological explanations; and whatever shortcomings such explanations might have are not traceable to the vacancy of the concepts attributed. So there seems to be nothing in the nature of a kind term that requires it to be non-empty, and consequently, there is no barrier to supposing that many non-empty kind terms now treated as world-dependent by the externalist might counterfactually be empty but meaningful.

Nor is it plausible, on Segal’s account, to suggest that referring terms and their non-referring counterparts express different concepts. To take an example,

Myalgic encephalomyelitis (ME, also known as “chronic fatigue syndrome” or “CFS”) is a condition characterized by chronic tendency to extreme fatigue ... At the time of writing, there is disagreement in the medical profession about whether ME is caused by a virus, and, if it is, how the process occurs ... Let us say that it is epistemically possible that there is no such thing as ME ... Not only is there no common underlying ailment, but there is also no reason to count ME as worthy of recognition as a syndrome ... But it is also possible that ME will become a natural-kind term ... "ME" could come to refer to the disease caused by the virus and nothing else ... Let TE1 be the empty case and TE2 the nonempty one. Let our twin subjects be Peter1 and Peter2. We must allow that Peter1 in TE1, expresses a genuine concept by his term "ME." If we do not, then we will have no adequate explanation of his words and deeds. Call this concept "C1." Let C2 be Peter’s concept and let "ME2" be our word for expressing C2. So "ME2" means just what "ME" actually means if our world happens to be TE2. And let C1 and C2 be individuated by their cognitive content. (2000: 37-41)

Segal now offers three arguments to show that C1 = C2. The first concerns the cognitive roles of C1 and C2. Suppose that the cognitive role of a given concept is defined by its causal role relative to actions and other cognitive states. Suppose further that a psychologist studying the two Peters assumes that C1 is identical to C2. According to Segal, this assumption is perfectly justified, because the role played by C1 in Peter1’s psychology is exactly the role played by C2 in Peter2’s psychology. Peter1 would reason and act in the same manner regardless of whether he had C1 or C2, and so would Peter2. Attributing either concept to either twin would, therefore, make no difference to the psychological explanation of the twins’ behaviour. And this, claims Segal, provides good reason for thinking that either attribution would be correct.
The next argument assumes that concepts have "supervenience bases." That is to say that if a subject S has concept C, there exists some set of physical properties intrinsic to S, and possibly relations between S and his or her environment, that are both necessary and sufficient for S's possession of C. Or, to put the point more roughly, there is a minimal set of physical properties and relations in virtue of which S has C. With that assumption in the background, Segal's argument runs as follows. It is allowed that Peter₁ has some concept C₁ expressed by his word "ME." Peter₁'s possession of C₁ has a specific supervenience base (base₁), so that there exists some set of physical properties and relations in virtue of which Peter₁ has C₁. If you duplicate base₁, you get a counterpart of Peter₁ who also has C₁. But base₁ is duplicated in TE₂, where Peter₂ is a counterpart of Peter₁. So Peter₂ has C₁. But if Peter₂ has C₁, then C₁ is expressed by his use of "ME." So C₁ = C₂.

The supervenience base constitutes the synchronic conditions sufficient for concept possession, but Segal also uses the diachronic (or developmental) conditions sufficient for the possession of a concept to give yet another argument that C₁ = C₂. The argument runs more or less as before. Consider how Peter₁ acquired his "ME" concept. We may say that he acquired it from his doctor, but Peter₁ was already in a position to rapidly acquire concepts simply by hearing a few words long before he had any symptoms or sought the advice of a medical professional. In order to explain this, we would have to construct a complex theory of his innate endowments and developmental history. The key to the present argument is that, whatever the nature of that theory might be, all the conditions specified by it would be present and active on TE₁; there is nothing available on TE₁ to explain how Peter₁ acquired C₁ (his empty concept of ME) that is not present on TE₂. But then Peter₁ has C₁, and C₁ is the concept he expresses by "ME." So the difference between C₁ and C₂ has to do at most with the extension conditions of the two concepts.

Aside from externalist arguments involving natural-kind concepts, Segal also attacks the kind of externalism based on sociolinguistic habits. Suppose TE₁ and TE₂ are identical in every respect except that on TE₁ the word "arthritis" refers exclusively to joint inflammation, whereas on TE₂ the same word refers to all rheumatoid diseases. Both Peter₁ (residing on TE₁) and Peter₂ (residing on TE₂) believe that one can get arthritis in the thigh. Recall that Burge used this example to argue that the content of one's thought depends on the linguistic conventions of the society to which one belongs. Deferring to the experts on their respective planets, the twins' concepts of arthritis (C₁ and C₂ respectively) would therefore differ.

But Segal rejects this reasoning. He believes that the identity of C₁ to C₂ still holds—especially in view of the fact that the two concepts have the same cognitive role in the psychology of their respective owners—and that, contrary to Burge, Peter₁'s concept of arthritis differs from the one sanctioned by the experts in his society. For consider: it seems clear that Peter₁ does not believe that he has an inflammation of joints in his thigh, knowing full well that his thigh is not a joint. But he does not positively believe that he does not suffer from arthritis in his thigh. Peter₁ therefore has two different concepts that he expresses by "arthritis" and "inflammation of the joints." But the experts on his planet have only one such concept, to which both expressions refer. So it cannot be that Peter₁ is able to deploy the same concept as the experts merely in virtue of his partial understanding and differential disposition. If he did, he would only have the unique concept denoted

---

41 Dr. Steven Savitt has pointed out (in correspondence) that this step begs the internalism-externalism question; I think this is correct. The externalist may well deny that Peter₁ bears the same relations to TE₁ (where there is no viral cause of ME) as Peter₂ bears to TE₂ (where ME is caused by a virus). So, insofar as the "supervenience bases" underlying C₁ and C₂ include agent-environment relations (and Segal supposes that they could), the externalist may deny that Peter₁'s supervenience base on TE₁ is duplicated on TE₂ for Peter₂.
by both expressions in the experts' vernacular. So either Peter, and the experts express different concepts by "arthritis," or by "inflammation of the joints," or both.

**Two-Factor Accounts**  Some have sought to take refuge in a two-factor theory, believing that cognitive content has a narrow as well as a broad dimension. Segal is not inclined toward this option, however. He is critical of all two-factor theories, especially the narrow functionalist and descriptive varieties. Functionalism is the thesis that psychological states can be identified with properties specified in terms of actual or potential causal relations. A standard way to develop this idea is as follows:

Take your favorite content-using psychological theory, e.g. common-sense or cognitive psychology or some combination of the two. The theory specifies causal relations among (i) inputs (ii) psychological states (iii) outputs. Leave your input and output terms in place, but remove all the psychological terms and replace them systematically with variables. Each variable now marks a state with a specific type of causal role, the role being specified purely in terms of causal relations to inputs, outputs and other states. Each variable thus marks the functional role associated with the psychological state picked out by the psychological term it replaces. (2000: 88-9)

Now suppose that $T$ is a psychological theory specifying (internal and external) causal relations among psychological states, inputs, and outputs. A two-factor account would reject the identification of functional states with psychological states attributed by $T$ and replace it with an identification of functional states with *narrow* contents. The idea is to accept that psychology is, on the face of it, externalist and would ascribe different contents to twins. It would say, for example, that Adam believes that water is good for plants, while Twin-Adam believes that water, is good for plants. However, $T$ would assign these two states the same functional role, the same pattern of potential causal transactions. Adam's water beliefs would be caused by, and would in turn also cause, exactly the same sorts of things as Twin-Adam's water beliefs. Thus a two-factor functionalist would take $T$ to be a wide psychological theory with wide content attribution, but would abstract from it a narrow content identified with causal role.

The problem with this view, according to Segal, is that the narrow content of a psychological state transcends the state’s tendency to interact with inputs, outputs, and other psychological states. This is because the nature of the state’s interactions depends also on various features of the cognitive system to which it belongs. Take two individuals who are identical in their beliefs, desires, hopes, fears, and so on. Both are confronted by a stampeding brown cow in a very bad mood. This input causes them to think, “There’s an enraged brown cow coming my way.” Both are very frightened. One decides to run and makes good his escape; the other, frozen in fear, does not move. Segal argues that the difference between the effects of the belief across the two subjects need not necessarily stem from any difference in the *content* of their psychological states. They both fear stampeding brown cows equally, but one’s freezing up mechanism just happens to be more sensitive than the other. We thus have a situation where the same content is associated with two different functional roles, and consequently, content cannot be functional role.

---

42 This is in fact a description of how to construct a “Ramsey Sentence.” See Block’s “Introduction: What is Functionalism?” for more details.

43 I doubt the functionalist would find this objection persuasive, since (s)he would not accept Segal’s separation of the content of a mental (i.e., functional) state from how the state is integrated with input conditions, behavioural interfaces, and other states in the system. So Segal begs the question against functionalism when he assumes that two functionally distinct states may have the same psychological content.
Two-factor theories can also be constructed by distinguishing the extensions of concept-terms from associated inner descriptions in an agent’s head. One might argue, for instance, that most people think of tigers as yellow-brown, black-striped, carnivorous, maneless, feline, dangerous, furry, etc. So perhaps the narrow content of one’s tiger concept is fixed by some such set of descriptions. But which one? Segal claims that there is no principled way of ascertaining which descriptions to include in the set. On the one hand, we have superficial descriptions that we can do without: albino tigers are neither yellow-brown nor black-striped, and a timid tiger, though unusual, is still a tiger. On the other hand, it is hard to make sense of a tiger that is not feline, or carnivorous, or mammalian. So some bits of description seem essential. The problem with these apparently essential descriptions, however, is that they do not suffice to fix cognitive role: we know too many species of feline carnivorous mammals.

The superficial features must therefore be included. But because we have no clear way of including some and excluding others, we must include them all. The result is a kind of sweeping holism that (directly or indirectly) connects my concept of tiger to every other concept I possess. Hence, it is a part of my tiger concept that tigers are physical objects, but of course physical object is a core part of numerous other concepts I have, each of which will contain features and descriptions that will be part of yet other concepts, and so on. The task now becomes to account for the cognitive content of a concept by reference to its relations to other concepts. So we are driven back to a kind of functionalist descriptive account that is very difficult to incorporate in psychology. For every time we wish to attribute content to a psychological state, we would have to specify the subject’s entire theory of the world.

**Radical Internalism** Segal maintains that externalism, in many of its current manifestations, is based on the common mistake of attaching too much significance to the intuitions generated by the kind of thought-experiments Putnam and Burge concocted. There is a tendency to think that Adam believes that water is good for plants while and Adam does not, that Peter believes that he has arthritis in the thigh while Peter does not. The error lies in taking this tendency too seriously, then drawing from it a general conclusion about the extension of concepts: either cognitive content is wide or it does not determine extension. The conclusion with respect to (common-sense) psychology then becomes: either psychology is externalist or it is in need of serious revision.

The idea Segal espouses is quite different. He insists that psychology, as it is practiced by layfolk and scientists, is already internalist. The externalist intuitions generated by the focal twin-worlds thought-experiments are simply misleading. They reveal only an accidental strand of our psychological thinking, but the basic apparatus of psychology does not mandate externalism. So ascriptions of content made by good psychology are already narrow, which is to say that they are intrinsic to agents and, therefore, equivalent across twins. Internalism, therefore, does not need to posit any additional notion of content beyond that which is already at work in psychology, that which is already attributed by content sentences of propositional attitude reports.

There is a caveat, however. Segal points out that if we are to correctly, precisely, and explicitly describe Adams’ and Peters’ minds, we would do well to adopt neologisms. We could say, for instance, that the Adam twins believe that “xwater” is good for plants, and the Peter twins believe that they have “xarthrits” in their thighs. The reason we need to adopt neologisms in these cases is not that the concepts we attribute are narrow; it is simply that our words “water” and “arthritis” do not adequately express the concepts of either set of twins. So if we are to use straightforward
propositional attributions correctly, precisely, and explicitly, we would have to employ new words. The use of neologisms, however, raises some obvious questions: Can we ever know what “xwater” and “yarthritis” mean? What is in the extension of these narrow concepts, and how can we find out?

Segal proposes that pre-scientific terms of natural phenomena be understood to apply to what he calls motleys (2000: 132). A motley may consist in several natural kinds, or in a collection that includes some, but not all, samples of a plurality of natural kinds. When a word becomes a natural-kind term with the development of science, its extension alters. Competent users come to regard some particular scientific principle of classification as correct, and so begin to use the term in line with that classification. The extension of the term may consequently enlarge, or shrink, or alter its boundaries in multiple directions. But according to Segal, before the scientific principle is known and explicitly adopted, there is nothing that ties the word to a unique natural kind.

It follows from this analysis that both the word “water” in 1750 and Adam’s xwater concept applied to a motley of objects. More specifically, in 1750 “water” was a term whose extension conditions did not confine it to any natural kind, but left open the possibility of its being true of many different objects. Had there been any XYZ in the universe, the word and the concept would have been true of it. Note that it is precisely here that neologisms may prove helpful. Segal predicts that neologisms are likely to be of use to psychologists studying subjects that differ in some important and general ways from the academics studying them—e.g., children, subjects from non-scientific cultures, and historical figures.

One step toward determining what is included in the extension of narrow concepts like xwater and yarthritis is simply to do folk psychology. We explain Adam’s behaviour by reference to his desires and beliefs, which, we presume, tend to cause what they tend to rationalize. This coordination of rationality and causality, holds Segal, lies at the heart of psychology, and offers us an obvious heuristic for ascribing content: charity. We can, and in fact do, get a lot right if we attribute to individuals psychological states that would render their behaviour rational.

But folk psychology is just the beginning. To learn more about concepts, we would eventually have to proceed to science. This transition need not widen the concepts under consideration, for scientific psychology exploits the same basic apparatus as do regular folk. Cognitive science, with its computational models, psycholinguistics, and complex of tacit theories, relies on something very much like the basic principle of charity. Segal thus believes that scientific psychology as it is actually practiced is perfectly compatible with internalism. So the right way to find out about narrow content is just the right way to find out about cognitive content in general: do psychology.

3.5 Descriptions, Motleys, and Neologisms

We have covered enough ground in this chapter to give a reasonably good idea of the motivations behind, and the various arguments in support of, both individualism and externalism. I want to conclude by saying a few words about where I stand on some key issues.

---

44 Segal defends the use of neologisms by first arguing that the proposal does not threaten us with widespread revision of current lay or scientific practice (2000: 142): the need to adopt neologisms only arises in cases where the ascriber lacks a ready word for a concept of the ascribee’s. Secondly, he observes that the use of neologisms already has some currency, and goes through a list of supporting examples (2000: 147-9).
One of the reasons I included the views of Segal in my exposition is that I have independently contemplated and developed a number of the notions he endorses. In particular, I find the idea of a narrow content compelling—in fact, inescapable—and I am also attracted by the functional identification of narrow content with actual and potential causal relations: the psychologist, I believe, has every right to consider mental states \( m \) and \( n \) content-identical, provided that \( m \) and \( n \) play the same functional role in their respective hosts—i.e., produce the same overt and covert behaviour, the same dispositions to act in a particular way, the same pattern of interactions with other psychological states, etc.\(^{45}\)

I also agree with Segal that narrow content is an intrinsic physical property of agents and, as such, the conditions of its specification and ascription do not depend on any external relational facts between agents and world. Both this and the previous consideration suffices to render radical externalism false: if narrow content is a real intrinsic property of agents, then it would be wrong to regard Adam and his Twin-Earthian doppelgänger as having different beliefs about the stuff they call “water,” just as it would be if we define narrow content in terms of functional role. But Segal offers another argument for the identity of the twins’ mental states: he claims that the contents of their beliefs—and the relevant (pre-scientific) terms used to describe these beliefs—are not confined to a single kind.

My own attitude concerning the move to assign motleys (i.e., an assortment of objects falling under one or another description) to the extensions of narrow contentful states is positive. Speaking of the directedness of an Intentional state involving an object \( o \), I stated that Intentional directedness is very much dependent on whether or not an agent is able to discriminate \( o \) from other objects, which, in turn, depends on the kind of information the agent receives from \( o \) and the manner in which this information is processed. In other words, whether or not a mental state is directed exclusively toward \( o \) depends on the resources of the information-processing system (agent) whose mental state it is, on whether or not these resources are sufficient to differentiate \( o \) from other similar objects. If not, then the mental state in question cannot be directed toward \( o \) exclusively. We can here identify the beginning of an argument-sketch for motley’s.

With respect to neologisms, it is likely that their extensive use will make some people very nervous, particularly those who are inclined to take propositional attitude reports at face value. On the other hand, I believe Putnam had it basically right when he decided “totally [to] abandon the idea that if there is a difference in meaning ... there must be a difference in our [narrow] concepts.” The remark suggests that Putnam has drawn a distinction between the narrow content of thought and the meaning of the words used to express these contents. One can take this to be an important first step in vindicating at least the occasional use of neologisms, but I shall remain neutral on this point. Segal encourages the employment of neologisms in particular cases, and even identifies how they are to be used and by whom. But whether or not neologisms can be exploited with profit is a matter of theoretical expediency, no more and no less. I doubt they have any inherent value, and will therefore neither urge nor discourage their application.

---

\(^{45}\) Recall Segal’s objection to functionalism: we cannot assume that \( m \) and \( n \) have different contents simply because they cause different behaviour, since the manner in which a psychological state generates action depends, not only on the state’s content, but also on various features of the cognitive system to which the state belongs. Whether or not this is a valid criticism is irrelevant to the present point. My claim is that where two mental states have (already) been found to have identical functional roles, the psychologist may reasonably consider them to comprise the same content. This is especially important in cases involving twins who are microstructurally identical, a situation that excludes the kind of objection made by Segal.
Notwithstanding the various points of contact, Segal and I diverge on a number of important matters. First, I am not at all impressed with the radicalism of his internalism. I neither think that narrow content is all there is to Intentionality, nor that psychology is wholly internalist. The classic stimulus-response paradigm for psychological investigation—a paradigm that has permeated every branch of the discipline ever since the days of Hermann von Ebbinghaus—\(^{46}\) is ample evidence that psychological methodology is not, and has never really been, solipsistic. Rather the opposite. Psychology has always recognized that the subjects it studies do not exist in isolation, that they interact with one another and with their environment.\(^ {47}\) One of the more basic assumptions of psychology is that there are discoverable patterns to such interactions; and the psychologist’s job (to put it very generally) is to shed some light on these patterns.

Any theory of Intentionality with a legitimate claim to plausibility must manifest the same attitude; it must at the very least acknowledge—if not directly explicate—the important role of the environment in mentality. This has been the motivation behind my aboutness-directedness distinction: I consider it an obvious truth that a great many Intentional states are literally caused by external sources and, to that extent, have a non-individualistic dimension that cannot be swept under the rug. For there are specific reasons why Intentional state \(m\) should be generated (or stimulated, or activated, or triggered, or brought into being) by object \(o\); and any theory that pretends to be complete must offer some account of this connection. I have already suggested that an explanation of such causal relations is to be found in the evolutionary history of creatures capable of Intentional thought, and will return again to the topic in chapter VII.

But though there is a wide dimension to Intentionality, the narrow features are theoretically richer and much more interesting. They certainly come much closer to representing what it is to be a conscious, self-aware being capable of thought and reason (or so it seems to me). Perhaps this is why Segal (among others) has been drawn to extreme versions of internalism, versions that take narrow content to be all there is to the matter. It might also explain why Segal rejects two-factor theories: the externalist dimension of Intentionality just doesn’t seem to be all that important. But this is wrong, for there is a difference between saying that directedness is more theoretically interesting, perhaps even more important, than aboutness, and saying that aboutness is neither interesting nor important at all. Aboutness certainly matters, though it does not dominate.

If I had to categorize the approach I favour, I would include it under the “two-factor” heading, with Intentional aboutness as the externalist factor and Intentional directedness as the internalist factor. Generally, I do not think Segal gives two-factor theories their due respect, and this is especially true of descriptive accounts. What is it that makes Adam and his twin react in exactly the same way to a pitcher full of water and a pitcher full of water? The answer cannot be a similarity in the chemical composition of the two substances, since one is made up of hydrogen and oxygen and the other is composed of XYZ. In fact, by stipulation, neither subject knows anything about the microstructural properties of the substance he sees. Rather, the obvious answer—and the correct one, in my opinion—must involve the macrophysical nature of the liquids. Both are colourless, odourless, tasteless, and non-viscous; each falls from the sky, runs in rivers, accumulates

\(^{46}\) One of the founding fathers of psychology, and the first to attempt to study human memory by objective means, using the classic “nonsense syllables” method; see Ashcraft (1994).

\(^{47}\) Well, perhaps not always. The selbst-beobachtung (or “introspective”) approach to psychological investigation favoured by Wilhelm Wundt and Edward Titchener (among others), suggests that there was a time when psychology was much more individualistic than it seems to be at present, or at least that some aspects of psychology’s past can be so described. Wundt’s and Titchener’s research interests and techniques were confined to “conscious processes and immediate experiences”—what psychologists would today identify as the areas of sensation, perception, and attention.
in chasms to form lakes, seas, and oceans; each is used to bathe, wash dishes and clothes, and "water" the lawn. At least some of these properties—or "bits of description," if you like—are sufficient for the twins' recognition of the contents of their respective pitchers. Just which properties fit the bill is a contextual matter. For now, it is important only to recognize that some description-based explanation seems to be mandated by the facts.

This brings us to Segal's main objection to descriptive accounts of content: they entail a sweeping holism that no feasible psychological theory can manage. I find this criticism unmoving for three reasons. First, I think that some measure of holism is necessary to explain certain mental realities: it is undeniable that my concept of tigers includes the idea of a solid, and it is also undeniable that the notion of solidity—perhaps a concept in its own right—enters into, bonds with, and constitutes a part of, many other concepts I possess. If these facts suggest holism, then holism it must be! But, secondly, I doubt that descriptive accounts entail the kind of extreme holism Segal attributes to them; for it also happens to be true that I can follow and understand a conversation about tigers with a perfect stranger: I know exactly what the stranger means by "tiger," even though I know nothing of him personally, much less his entire theory of the world. Finally, it should not matter even if descriptive accounts turn out to be acutely holistic. While it is not inappropriate for philosophers to engage in the business of plotting the practical boundaries of psychology, it is a serious mistake to allow these boundaries to dictate the course of philosophical theorizing. Philosophers ought to be interested in the facts wherever they may lead. And if, finally, these facts prove too complicated for current psychological apparatus, then psychology had better make haste and widen its stride.

To summarize: I think that Intentionality has both an externalistic and an individualistic component, and so a two-factor theory is the proper route to take. The external dimension of Intentionality—or Intentional aboutness, as I have called it—connects the environment to narrow contents by causal relations, the explanation of which is to be found in the evolutionary history of humans. The internalist dimension of Intentionality—what I have called Intentional directedness—manifests the following features:

a. Narrow content is an intrinsic physical property of subjects and is neither specified nor individuated by means of external relations between subject and environment.

b. Assuming that functional role supervenes on microphysical structure (and I assume it does), we can identify narrow content with functional role; in which case, sameness of microphysical properties would entail sameness of function role, and sameness of functional role would entail sameness of narrow content (so sameness of microphysical properties means sameness of narrow content).

c. For systems that lack perfect discrimination (and that covers just about everything, including humans), narrow contentful states will not be confined to specific kinds, and this leaves open the possibility of using neologisms as a theoretical aid.

d. As people seem to recognize objects and judge situations mostly by sensory cues, our account of narrow content must involve some kind of descriptive story. This might suggest some measure of holism, which, far from being a disadvantage, is actually in line with how the human cognitive system seems to be organized.

e. Finally, our account of narrow content must explain its aspectual shape and role in error-generation.
There it is! The above five points constitute the desiderata for any acceptable theory of Intentional directedness. The first of these—the claim that narrow content is an intrinsic property of brains—is the key to the rest; but ironically, it has received very little attention. Segal, for instance, does little more than assert (a) at various places in his narrative. But to sketch out the physical foundation of Intentional directedness, and to account for the other desiderata besides, is to give the whole notion of narrow content a legitimacy that, I fear, may be presently lacking: we would be moving from assertion to demonstration, from merely saying that narrow content exists to showing how it can be physically implemented. To that end, scripts should prove very valuable.

I do not offer scripts as an ultimate or complete theory of Intentionality, nor do I think that any one account, or single kind of explanation, is likely to be so exhaustive. The vast complexity of the human brain seems to suggest otherwise. But I am reasonably confident that scripts do have a place in a general explanatory account of Intentionality. The aim of the rest of the dissertation is to notch out that place.
Scripts I: Foundation and Structure

And a Word About the Game

There is an universal tendency among mankind to conceive all beings like themselves, and to transfer to every object, those qualities, with which they are familiarly acquainted, and of which they are intimately conscious. We find human faces in the moon, armies in the clouds...

David Hume, *The Natural History of Religion*, p. 117.

Sometimes we see a cloud that's dragonish;
A vapour sometime like a bear or lion,
A tower'd citadel, a pendant rock,
A forked mountain, or blue promontory
With trees upon't, that nod unto the world,
And mock our eyes with air...

William Shakespeare, *Antony and Cleopatra*, IV.XII.

4.1 The Game

I grew up in a part of the world that, at the time of my early childhood, had not yet been infiltrated by modern technology. Summer vacations were a prolonged exercise in imagination: there were no computers or video games; there were no organized sports (except for soccer, which was usurped by older teens and young adults); children’s television programming was scarce and radio entertainment was the province of grouchy old men who loitered in coffee shops and reminisced about times long past. In short, the environment in which I was raised did not furnish many possibilities of amusement beyond wide open spaces and a number of neighbourhood kids who, like myself, were always looking for something to do. We had to create our own fun, and so we did, with all the vigour and determination that can be afforded by the will of idle preteen children.

We invented many pastimes, but one particular game stands out from the rest. It was a favourite among our group, and one which—in order to ensure maximum participation—did not impose a limit on the number of players. The game required a bare cement wall, a small container of some kind (usually a plastic cup), and a nearby source of water. Each player had to fill the cup with water which he would then splash against the wall. The action produced a dark water-stain that had

---

1 The players in my group were almost always other boys. Girls were simply not allowed, except in those rare circumstances when there were too few of us to carry out an effective game, or when the only cup that can be found belonged to a certain girl who, in exchange for the use of her property, demanded full and equal participation. But even then, girls were admitted against the better judgement (and loud protestations) of almost every member. Of course, girls formed their own cliques, which, quite naturally, did not admit members of the male persuasion.
the tendency to evaporate quickly. The object of the game was to try to determine what the stain was. It might have been an animal, or an ordinary item like a chair or a tree. It might have been an insect or a person. Teachers, mutual friends, and even family members were sometimes seen in those wet patches.

Such variety notwithstanding, a player was not at liberty to name objects simply at random. There had to have been a sufficient measure of resemblance between the shape of the water-stain and the object named. The entire group was to judge on what exactly constitutes a “sufficient measure of resemblance” for each individual case. Politically motivated refereeing was not common, since every judge had to be judged in turn, and those who might have been unfairly treated were not above taking hasty revenge. Each player was given two chances to name the “object” on the wall; if he could not, he was disqualified, and the game would continue in this way until there was a final winner.

Thinking back, I am struck by the number and variety of objects we were able to find: everything from chimps, daisies, and seashells to spaceships, monsters, and religious icons. That is not to say that there were not favourite items. Elephants, donkeys, and spiders were particularly common motifs. No less remarkable was the clarity with which we saw the objects. If someone identified a horse, for instance, at first we may have been unable to see it. But once the form of the horse had been outlined, a curious transformation took place. The stain on the wall was no longer a haphazard, jagged figure. It had suddenly become a horse. It had always been a horse. And try as we may, we were never able to re-establish the initial quality of irregularity the stain seemed to display only moments earlier. For this reason, there was seldom divergence of opinion about the items found. On the other hand, agreement was neither insincere nor capricious. Verisimilitude—to the extent that we perceived it to be there—was the order of the day, and when we saw none, sanctions quickly followed.

I remember myself to have been an above average player (biased as this assessment must be). While there were others who were more proficient, most of my playmates looked upon my talent with some trepidation. But the truth is, rank was not all that important to anyone (though no one would admit as much). We cared mostly about the immediate gratification of having fun and being in each other’s company.

The objects we saw in the water-stains, though fascinating in their own right, are not of immediate concern. What is of interest is how we came to see those objects. Any objective comparison of the wet patches and the things they supposedly resembled would surely reveal only the barest similarity between the two, certainly not enough to make the former adequate images of the latter. Yet the many agreements reached by the players about the identity of the shapes, and the speed and ease with which those decisions were normally made, seem to resist this conclusion. Unless one is disposed to overlook the whole exercise as arbitrary, one must allow that, however clumsy a water-stain may have been, it must have displayed enough resemblance to the “real world” object to make the subsequent identification not only possible, but unanimously sanctioned. The question is, how was this identification accomplished? What is it that makes us see human faces in the clouds, armies in the sky?

---

2 The number of chances given to each player was largely influenced by certain climatic realities of the region: unless the identity of the water-stain was ascertained within the time limit of two quick guesses, the stain would evaporate and disappear under the hot sun. Players usually tried to avoid this since, as a general rule, the level of skill of a member was inversely proportional to the number of times he has “caused” an evaporation.
Whatever one may say about the matter, one thing seems obvious: from a layperson’s point of view, there is nothing remarkable going on here. An individual’s visual field is a virtual gallery of shifting colours, shades, forms, and movements. Some of these are easily identifiable as a rosebush, a passing automobile, a hotdog stand, or the neighbour next door. Other times, when direct perception is obstructed, we lapse into a shakier trial-and-error strategy. The water-stain game is essentially a variation on this theme. It is merely a novel and amusing way of performing a task each one of us carries out almost every waking moment.

From the perspective of psychology, however, the mechanisms which enable an individual to recognize and identify a thing have generated enormous interest. What I wish to do in this chapter is to describe a theory that provides some interesting insights into the matter. Script theory, as it is called, is fundamentally a hypothesis about how information is organized and stored in memory. As such, its implications go far beyond the simple recognition and identification of things.

Scripts exploit an important fact about the human brain—namely, the brain’s extraordinary ability to model the environment of which it is a part. In this modelling lies, not only our capacity to recognize things, but also our ability to understand, think, and undertake almost every higher-order cognitive function. And in this modelling too, I want to argue, is found the very heart of Intentionality. I am not alone in holding this view. McGinn believes that the Intentionality of a mental state has a logical shape that derives from the brain’s ability to model, which occurs in something like “an engineer’s workshop in which no one speaks,” but engages in “the production of vastly many practical models” (1989: 208). The phrase “no one speaks” suggests that these models are non-linguistic; the phrase “much more than ordinary geography” suggests that some of their components may be sensitive to the logical space into which objects and events are placed.

More than half a century ago, the Scottish psychologist Kenneth Craik took the brain’s modelling capacity as a central topic in his one and only published book, *The Nature of Explanation*. Craik assumed that any reasonably persuasive theory of rational action must adopt the modelling hypothesis as a starting point. His main idea was that human thought—or at least a good portion of it—has a kinship to the activities of machines that represent aspects of the world and then make predictions on the basis of these representations. Thus an engineer might use a special-purpose computer to predict the behaviour of a suspension bridge given the values of certain factors: the various weight thresholds of the suspension cables, the type and strength of the alloys used in the construction, the classification of the soil upon which the structure’s foundation will rest, the length and cross-sectional area of the supporting beams, seasonal variations in temperature, etc. Based on these values, the computer might then assist the engineer in working out the safest and most practical design for the bridge.

Craik believed that this kind of modelling is at the core of human mental life. “To those in sympathy with this attitude,” says he, “there is something wonderful in the idea that man’s brain is the greatest machine of all, imitating within its tiny network events happening in the most distant stars, predicting their appearances with accuracy, and finding in this power of successful prediction and communication the ultimate feature of consciousness” (1943: 99).

In a manner similar to the engineer’s computer, we make decisions by imagining what would happen if we do such-and-such, and then construct various scenarios in our minds about how others might react to our actions. Based on this information, we come to some conclusion about what ought to be done. Or sometimes we remember what occurred in similar circumstances to someone we know and then use that information to decide how to act. As Craik summed up in the closing paragraph of his book:
Assuming then the existence of the external world I have outlined a symbolic theory of thought, in which the nervous system is viewed as a calculating machine capable of modelling or paralleling external events, and have suggested that this process of paralleling is the basic feature of thought and of explanation. The possessor of a nervous system is thus able to anticipate events instead of making invariable empirical trial. (1943: 120-1)

Anticipating events without having to make "invariable empirical trial" is the best reason in favour of the modelling hypothesis. Craik's attitude toward the models in question seems to be a little more concrete than McGinn's, and so is mine. I propose that one way to think of these models is in terms of scripts. I will eventually argue that scripts are real neural structures that have been shaped—by evolution as well as the local learning history of the organism—to perform certain indicatory and indentificatory tasks, and therein lies their content: the ability of the human brain to represent the environment is essentially its ability to (re)shape neural networks into becoming selectively sensitive to specific stimuli ranging over all sensory modalities.

Most of this chapter, however, is an introduction to script theory, not as it applies to the human nervous system, but as it has been developed by artificial intelligence researchers. Starting in §4.3, I will venture beyond Schank and Abelson's basic exposition by introducing four different types of scripts, only one of which seems to include linguistically-based information. I take these four kinds of scripts to be operative and widely utilized in human cognitive processing. How these scripts are physically implemented in the brain and whence they get their content are questions dealt with in chapter VI. The main burden of the present chapter is to introduce the central notions related to script theory in preparation of their later application to these questions.

4.2 Conceptual Dependency

The best account of script theory was given by Roger Schank and Robert Abelson in *Scripts, Plans, Goals, and Understanding*. I follow it closely in the expository sections below. But first, a brief word on the research methods of AI and psychology.

1. The Research Methods of AI and Psychology

*Scripts, Plans, Goals, and Understanding* is a work at the intersection of psychology and artificial intelligence (AI). "The psychologist who studies 'knowledge systems' wants to know how concepts are structured in the human mind," claim Schank and Abelson,

how such concepts develop, and how they are used in understanding and behaviour. The artificial intelligence researcher wants to know how to program a computer so that it can understand and interact with the outside world. The two orientations intersect when the psychologist and the computer scientist agree that the best way to approach the problem of building an intelligent machine is to emulate the human conceptual mechanisms ... (1977: 1)

There is, however, some difference between the methodologies of the two disciplines. North-American experimental psychology was dominated by behaviourism for so long that the study of cognitive process lay almost dormant, while other branches of psychology (developmental, clinical, social, and psychophysiology) continued to evolve. By the mid 1940s, the introspective approach to psychological research had all but disappeared, and the results achieved in this way were no longer considered credible. The quantitative method upon which present day psychologists rely was developed in reaction to the growing dissatisfaction with the introspectivism that saturated early
psychology. Eventually, however, behaviourism declined, and the nature of mental phenomena started to re-attract attention. Still, the echoes of old suspicions die hard: even today, acceptable scientific procedure in cognitive psychology calls for quantitative response measurements—such as accuracy of recall, choice, or reaction time—when subjects are confronted with stimulus tasks.

In comparison to psychology, AI is relatively new. Its early efforts were primarily directed toward getting computers to solve logico-mathematical problems and play games like chess and checkers. Nevertheless, it had always been considered important that computers be made to deal well with natural language (Schank and Abelson 1977: 8). The problem is that natural language comprehension is an impressively difficult skill, for there is no clear way to understanding natural language without at least a remedial ability to manipulate concepts or to model certain aspects of the (social and physical) world. Accordingly, a good deal of AI research now involves speculation about how the required modelling is to be effected. Researchers theorize about what might be involved in carrying out certain cognitive tasks and then program their models on computers to reveal deficiencies or errors in their theories. There are no experiments, at least not in the manner performed in cognitive psychology.

A very important development in AI research over the past twenty years has been an increasing recognition that context is of prime importance in the interpretation of natural language strings. Implicit real-world knowledge is applied by the understander, with the effect that clues for extracting the role of a certain concept are often nowhere to be found in the strings themselves. One example is,

(1) The policeman held up his hand and stopped the car.

Most of us understand this sentence effortlessly by imagining a driver who, in response to seeing the policeman’s raised hand, steps on the brakes, gradually slowing the vehicle down before coming to a complete stop. None of these intermediate steps is mentioned in the sentence, however. Nor does there seem to be any causal connection between a policeman’s raising his hand and the stopping of a car, outside of specific social conventions that are also not mentioned in the sentence.

(2) I am fond of pickled octopus.

The speaker is talking about eating pickled octopus, but this is not directly stated. Nor should it be. Unless the speaker has good reason to be explicit about a certain piece of information, it is practically necessary that (s)he assume that the audience is capable of inferring a great many details for themselves. This is what makes speech a two way process, for we would find talking very arduous if we had to make everything to which we refer explicit.

The above two examples should demonstrate just how difficult it can be to understand natural language, much less model it on a computer. Of course, we find it easy to communicate because we take for granted all the background knowledge our audience brings to a conversation. But the AI researcher has no such luxury: (s)he is forced to specify each and every step that accounts for a program’s ability to utilize even the simplest information. So the extent and manner in which background knowledge aids comprehension become the central issues. It is here that AI can use help from cognitive psychology. But the current psychological demands for quantitative measurements—while by no means trivial—may be too restrictive for the job at hand: experimental tasks are often much different from those encountered in daily life, a disparity that has led R. J. Spiro to remark,
In general, the subject can be expected to assume that the information in the discourse is of no future usefulness. The discourses are typically and clearly fictional. Even if they are perceived as true, their truth and any other considerations regarding the topical content vis-à-vis prior knowledge are irrelevant to the purposes of the experiment. The usefulness of the discourses begins and ends with the experiment. One of the main reasons in everyday life for relating new information to old is negated: selectively processing information in order to update one’s knowledge (that is, keeping the knowledge "current") of issues which are personally interesting or important. It would be foolish to update one’s knowledge with the useless, isolated, and probably false information usually found in experimental prose. (1977:140)

Spiro is doubtful about the ecological validity of the typical research methods of psychology—especially when it comes to investigating our representations of ordinary situations. Though it may seem that neither the procedures of psychology nor those of AI are independently adequate for investigating human knowledge structures, no broad compromise between the two fields was ever forged. Both psychology and AI followed their own pursuits, each keeping an eye on the advances of the other: AI continues to view the computer as “an omnipotent, but very dull and plodding, god,” (Schank and Abelson 1977: 20) while cognitive psychology continues to rely on stimulus-response tasks as a basic research paradigm.

ii. The Axioms of CD

We cannot speak of scripts without talking first about Conceptual Dependency (CD) theory, since the former is the offspring of the latter. CD was introduced by Roger Schank as a theory of the representation of the meaning of sentences. At the time, computer programs were already parsing words into various grammatical categories. Conceptual Dependency took the research one step further. It was the first sophisticated attempt to provide a computer with the kind of resources necessary for processing natural language sentences and their interrelations. The basic axioms of CD are (Schank and Abelson 1977: 11-12):

[A] For any two sentences that are identical in meaning, regardless of language, there should be only one representation.

[B] Any information that is implicit but still necessary to the understanding of a sentence, must be made explicit in the representation of the meaning of that sentence.\(^3\)

[C] The meaning propositions underlying language are called conceptualizations. A conceptualization can be active or non-active.

[D] An active representation has the form: Actor, Action, Object, Direction (instrument: inst).

[E] A non-active conceptualization has the form: Object (is in), State (with value).

The first of these axioms directs us to look closely at sentences that seem to convey the same information and to somehow extract the core of their similarity. The second forces us to make explicit whatever differences there might be between two semantic units and to express these differences accordingly. Thus two verbs in a language may share the same primitive element—e.g., “walk” and “run” share the primitive element change of location—but may also have differences. In

\(^3\) There are two assumptions here: (1) at least some of the information implicit in a sentence might be unnecessary for understanding that sentence; and (2) for every sentence, \(T\), the amount of information necessary for understanding \(T\) is finite. If these assumptions are correct, then [B] need not lead to any sort of infinite regress.
such cases, the proper representation of the verb will be the primitive element it shares with other verbs, plus the explicitly stated concepts that make it unique.\textsuperscript{4} [C] is suggested by [A]: if there can be identity of meaning across varying manners of expression (i.e., across different natural languages), then it might be helpful to separate this underlying meaning from the strings of symbols that communicate it: hence, "conceptualizations."\textsuperscript{5} The last two axioms describe the two possible states of these conceptualizations.

iii. Primitive Acts

Because an active conceptualization is defined as an actor doing something to an object in a certain direction (or by means of some instrument), it will be necessary to determine what an actor can do. The primitive acts of Conceptual Dependency are (Schank and Abelson 1977: 12-14):

- **ATRANS**: A transfer of an abstract relationship such as possession, ownership, or control. Thus, one sense of "give" is: ATRANS something to someone else; one sense of "take" is: ATRANS something to oneself; "buy" comprises two conceptualizations that cause each other: one is an ATRANS of money, the other an ATRANS of the object being bought.

- **PTRANS**: The change of physical location of an object. Thus, to "go" is to PTRANS oneself to a place; to "put" is to PTRANS an object to a place. A PTRANS always results in a change of location (ALOC).

- **PROPEL**: The application of a physical force to an object. PROPEL is used wherever any physical force is applied, regardless of whether or not the force results in the movement (PTRANS) of the object. "Push," "pull," "throw," and "kick" have PROPEL as part of them. "Samson destroyed the two main pillars" is a PROPEL that causes a (catastrophic) PTRANS. "Adam threw the stick" is a PROPEL that involves a GRASP (see below).

- **MOVE**: The movement of a body part of an animal by that animal. MOVE is often an act in an instrumental conceptualization for other acts. That is, in order to throw, it is necessary to MOVE one's arm; and MOVE foot is the instrument of "kick."

\textsuperscript{4} Schank elaborates on this point in §4.2 of his (1972).

\textsuperscript{5} This is consistent with Schank's pre-theoretical stance expressed in his initial exposition of CD: "The theory presented here has as its initial premise that the basis of natural language is conceptual. That is, I claim that there exists a conceptual base that is interlingual, onto which linguistic structures in a given language map during the understanding process and out of which such structures are created during generation" (1972: 553-4). Riesbeck expresses the same sentiment: "Basic to Conceptual Dependency theory is the assumption that thoughts are different than sentences, that ideas are built from language free concepts and language free relationships between those concepts. Communication is the transformation from a thought to an utterance by one person, back to something like the original thought by another person. The transformations are between structures of two very different types, between concepts and sound patterns" (1975: 100). Compare this to Stalnaker's (1970: 33) notion of *proposition*: "Propositions are things that may be considered in abstraction on the one hand from particular languages and linguistic formulations (the sentences that express them), and on the other hand from the kinds of linguistic acts in which they figure (for example the assertions and commands in which a proposition is asserted or commanded). Dretske (1988a: 33) makes a similar point in regard to content: "... my automobile's gas tank gets filled because I produce sounds with a certain meaning. I say 'Fill it up.' If I produce sounds with a substantially different meaning, the tank doesn't get filled. And if, at a different place and time, I produce completely different sounds with a similar meaning (e.g., "Benzina, per favore"), the same result is achieved. So it looks like it is the meaning of the words I produce (i.e., what I say, not the sounds I produce in saying it) that is having the desired effect."
GRASP The grasping of an object by an animal. The verbs "hold," "grab," "let go," and "throw" all involve the act of GRASPing.

INGEST The taking in of an object by an animal to the inside of that animal. "Eat," "drink," "smoke," and "breathe" are common examples of INGEST.⁶

EXPEL The expulsion of an object from the body of an animal into the physical world. Whatever is EXPELED is very likely to have been previously INGESTed. Words of excretion and secretion are described by EXPEL—among them "sweat" and "spit."

MTRANS The transfer of mental information between animals or within an animal. Schank and Abelson partition memory into two types: the CP (conscious processor, where something is thought of), and the LTM (long-term memory, where things are stored). The various sense organs can serve as the vehicles of MTRANS. Thus "tell" is an MTRANS between people that involves SPEAK (see below); "see" is an MTRANS from eyes to CP; "learn" is the MTRANSing of information to LTM.

MBUILD The construction by an animal of new information from old information. "Decide," "conclude," "imagine," "consider," are all examples of MBUILD.

SPEAK The action of producing sounds. Many animals can SPEAK, but only humans do so in order to MTRANS. The words "say," "whistle," "bark," and "thump" are examples of SPEAK.

ATTEND The action of focusing a sense organ toward a stimulus. ATTEND ear is "listen;" ATTEND eye is "look." ATTEND is almost always referred to in English as the instrument (inst) of MTRANS. Accordingly, in Conceptual Dependency, "see" is treated as MTRANS to CP from eye by instrument of ATTEND eye to object.

DO In CD, DO represents a primitive, non-specific act. It is used when the manner in which a given action is achieved is unclear. For example, to say "Samson killed himself" is to say that Samson did something to himself which resulted in a fatal state of health. What he did exactly is irrelevant, both to the representation of the English sentence and to its meaning.

Shank and Abelson admit that there has been much debate over whether these primitives are of the "right" sort, with some accusations that CD is ad hoc. They stress, however, that the important criterion is whether or not Conceptual Dependency is useful, especially when the aim is to organize information so that a computer can use it to effect natural linguistic interaction with humans. So whether or not the "right" sort of primitives are utilized is a question that must yield to expediency and to the outcome of subsequent computer simulations. More importantly, however, the "ad hoc" argument can be made for any list: it is always possible for someone to come along and say, "but you have missed such-and-such," or "you have improperly included so-and-so." If we take all of these claims seriously, we will never get off the ground. At some point we must (and should) ignore the critics and get on with our work. Be that as it may, Schank and Abelson do believe that any proposed replacement of CD will more than likely end up with the same primitives, and the systematic linguistic exploration of candidates for primitives conducted by Jackendoff (1976) seems to point in that direction.

⁶ The conceptual apparatus which enters into INGEST is outlined in detail by Goldman (1975: 330-2).
iv. Numeric Values and Scales

Conceptualizations expressible by attribute-value statements make use of a large number of scales. These scales run between -10 and +10. Schank and Abelson do not undertake a serious quantitative scaling of relative points along this continuum, but use numerical references only suggestively. Some of the scales they use, along with their boundaries and increments, are indicated below.

| HEALTH | dead, diseased, under the weather, tolerable, healthy. |
| ANTICATION | terrified, nervous, hoping, confident. |
| MENTAL STATE | devastated, depressed, neutral, happy, ecstatic. |
| PHYSICAL STATE | non-existent, damaged, neutral, perfect. |
| AWARENESS | unconscious, asleep, awake, keen. |

v. Representing Causation

We have stated the primitive acts of Conceptual Dependency and the conceptualizations that involve the use of scales. But determining what an actor can do is not enough without a way of conveying the fact that the actor was able to cause some kind of change. So representing causation is the next step. Again, context is very important here, since it frequently happens that the causal elements which connect actors, objects, and events together are not openly stated in natural language strings. Consider the following:

(3) Adam cried because Eva said she loved Samson.

Why did Adam cry? The literal meaning of (3) is odd, for certainly Adam did not cry because of the event of Eva’s speaking. Most readers know precisely why he cried: Eva does not love him, because she loves Samson and, typically, a woman does not love two men simultaneously in exactly the same way; but Adam loves Eva and (therefore) he wants her to return his affection; lopsided regard is not an agreeable state of affairs and often causes anger, violence, depression, etc. ... These are just some of the concepts—all of which implicit—that enter into our understanding (3). Here is another example.

(4) Samson’s leg was broken because Delilah tripped him.

(5) Samson’s leg was broken because Delilah knocked over a pile of bricks.

We are more or less comfortable with the causality in (4), since we know that tripping someone can cause a broken leg. But in sentence (5) there is no such direct connection, so we must infer the real causal event: the knocking over must have propelled the bricks in contact with Samson’s leg. This inference is necessary in order to understand the sentence, and demonstrates that causal connections cannot be taken at face value. If we hear that x caused y, we must ask if x could have caused y

7 Many more scales are covered by Goldman (1975: 341-5).
directly; if it could not, we must figure out the intermediate links. This is what Schank and Abelson call *causal chaining*.

The symbol “↑” denotes causality in CD, as in the following representation (the direction of the arrow is from caused to causer):

\[ \text{(6) Eva killed Adam} \quad \text{Eva DO} \quad \uparrow \quad \text{Adam HEALTH (-10)} \]

Schank and Abelson (1977: 25-8) posit the following causal axioms:

- [C1] Actions can result in state changes (↑r).
- [C2] States can enable actions (↑E).
- [C3] States can disable actions (↑dE).
- [C4] States (or acts) can initiate mental states (↑I).
- [C5] Mental states can be reasons for actions (↑R).

Not any action can result in any state, and not any state can enable any action. Thus every primitive act in CD is associated with a set of states that it can affect and a set of states that are necessary in order to affect it. This kind of delimitation takes Conceptual Dependency some way toward solving the context (or inference) problem. To see an instance of how this might work, let us return to sentence (5). There we find one action given—Delilah PROPEL (bricks)—and one state change, Samson’s leg PHYSICAL.ST(-). For PROPEL, the applicable rules are:

\( \alpha. \) PROPEL can result in physical contact (PHYSCONT) between the object of the PROPEL and any other object in its trajectory.

\( \beta. \) PROPEL results in PHYS.ST(-) if one of the objects in the PROPEL is human, if the PROPEL results in PHYSCONT with that human, and if the force of the PROPEL is great.

These rules, when applied to “Delilah PROPEL Delilah to bricks,” yield that Delilah is in PHYSCONT with the bricks, and according to \( \beta \), Delilah could be damaged by this action but not Samson’s leg. It is therefore necessary to hypothesize an action that could have resulted in a PHYS.ST(-) for Samson’s leg. Using the above rules backwards—i.e., \( \beta \) first then \( \alpha \)—we get the following hypothetical event:

Something\(_1\) PROPEL something\(_2\) to leg (Samson).

This hypothetical event must reflect some real-world occurrence if the stated causation is to make sense. This can be done with the help of another rule:

\( \chi. \) If a moveable object, \( d \), is put in PHYSCONT with another moveable object, \( b \), with sufficient force, then \( d \) can cause a PROPEL of \( b \), in which case, \( b \) may be put in PHYSCONT with a third object, \( s \), which may cause a PROPEL of \( s \), in which case, \( s \) may be put in PHYSCONT with a fourth object ...
Substituting Delilah for “d,” the brick for “b,” and Samson for “s”—together with rules $\beta$ and $\alpha$—it is simple to hypothesize as an inference,

Delilah PROPEL bricks to leg (Samson).

Thus we have what appears to be a way of generating correct inferences, once we know the set of states a primitive act can influence and the set of states with which it can intermingle.

Let us now consider a more complex example, which should bring together some of the different ideas discussed so far (Schank and Abelson 1977: 28).

(7) Adam was thirsty. He opened a can of beer and went into the den. There he saw a new chair, on which he sat down. Suddenly the chair tilted over and Adam fell on the floor. His beer spilled all over the chair. When his wife heard the noise, she ran into the room. She was very angry that her new chair was stained.

Figure 4.1 is a representation of the above story and the causalities implicit therein.

4.3 Scripts: Episodic, Instrumental, Personal

Schank and Abelson believe that it is the network of causal links that makes discourse coherent. Once the actual events that took place in a story are determined, a deeper level of data processing is possible. The problem is how to make explicit those elements that have been left out. Conceptual Dependency was designed to handle that problem at the single thought or sentence level. Causal chaining handles the problem at the level of interconnected thoughts or texts. Yet it is easy to see that CD is not powerful enough to process the kind of natural language strings that rely on causal connections or background information buried deep inside human social rules and cultural norms. For example, more is needed to represent sentences like,

(8) Eva liked Adam, but not enough to meet him at the Blue Moon Motel.

(9) Samson worked hard, got rich, paid a hefty dowry, and married the beautiful Delilah.

The human agent brings a large repertoire of knowledge structures to almost every aspect of cognition—perceptual information processing, object recognition, language comprehension, problem solving, decision making, and imagination—and it is extremely difficult to determine just what background information enters into a given cognitive skill.

The knowledge structures in question have sometimes been called schemata. Rumelhart (1977: 266) defines a schema as,

an abstract representation of a generic concept for an object, event, or situation. Internally, a schema consists of a network of interrelationships among the major constituents of the situation represented by the schema. Moreover, a schema is said to account for any situation that can be considered an instance of the general concept it represents.

Rumelhart uses this definition to construct a theory of understanding in which the process of comprehension is taken to be identical to the process of selecting and verifying conceptual schemata to account for the situation (or text) to be understood. In other words, “on having selected and
verified a schema and determined that it has given a satisfactory account of the situation in question, one can be said to have understood the situation or passage" (1977: 268).

Figure 4.1: A representation of the explicit and implicit causal connections in story (7). The representation says essentially this: Adam's thirst caused him to decide (MBUILD) to DO an action that resulted in a beer being opened, and this enabled him to drink (INGEST) it. The opened beer was at least partly responsible for the chair's being WET. This happened when gravity PROPELed the beer and caused it to be in physical contact (PHYSCONT) with the chair. The opening of the beer, however, has no causal connection to Adam's entering the den. So the latter action initiates an independent causal chain. Adam walks (PTRANS) toward the den, enters it (ALOC in den), notices the chair (ATTEND, inst eye), walks toward it (PTRANS), and sits in it (PHYSCONT). This causes a change in the chair's state, which partly enables one action and results in another. The action partly enabled by the change in the chair's state is gravity's PROPELing the beer onto the chair. The action that resulted from the change in the state of the chair is a PROPELing of Adam to the floor. Adam's accident causes a noise, which is heard by Adam's wife (ATTEND noise, inst ear, and MTRANS noise to CP), who becomes a little worried (ANTICIPATION [-5]), walks over to the room (PTRANS, ALOC in den), sees the chair (ATTEND chair, inst eye), which, together with the chair's being WET, eventually cause her to be angry.

The idea of a schema in the interpretation of human events has a long history in social psychology, which had its roots in Gestalt psychology. Founded by Max Wertheimer, Gestalt psychology surfaced as a theoretical school in Germany early in the 20th century. The movement is based on the belief that the whole is greater than the sum of its parts. The main exponents of Gestalt psychology maintain that psychological phenomena could be understood only when viewed as organized, structured wholes, not when broken down into primitive elements. Accordingly, this
branch of psychology focuses on ways to unite the mind and body in order to make a person “a single unit” (Ashcraft 1994: 579-86).

Since the incarnation of the Gestalt principle, a long line of social psychologists have appealed to structured ideations of the way people suppose the world to be organized. Script theory is a relatively recent addition to the list.

A word of warning before we go on. A bias toward linguistic entities like *sentences* and *stories* may have been noticed in the above exposition of CD, and a similar bias may also be present in the exposition of script theory to follow. But such partiality is restricted to pedagogy and does not reflect any linguistic orientation of the theories themselves. Because CD and scripts have been developed largely within the context of AI, most of the ideas involved are explained using computer exercises and demonstrations based on written passages. But such is a procedural prejudice, not a theoretical one.

Script theory is not limited to linguistic information. I have said that scripts should be understood as schemata, but in some instances the term “concepts” is also apt. Scripts have originally been developed as a hypothesis about the structure of memory, and as such, we shall find that different scripts vary widely in abstractness and representational mode. For instance,

a. Adam loves the Star Wars theme. He has what we may call an *acoustic script* of that particular musical piece. This script—residing in his auditory memory—allows him to determine things like, how far into the theme a department store rendition has gone and whether or not his favourite part has passed. The script also allows Adam to anticipate certain sequences and, generally, to make a variety of inferences related to the theme.

b. Samson has a concept of Delilah and that concept is also a script. What sort of information does Samson’s concept-script contain? The most basic information would be what Delilah looks like: a woman in her mid-twenties, blond, blue-eyed, about 125 lbs, etc. A lot of other details will be there as well. Delilah’s personality or character, what she likes and dislikes, certain autobiographical facts, and so on.

c. Eva has a restaurant script which tells her the sequences of events that take place in a typical “dining out” experience. The script also tells her who the players are, what they are supposed to do, and what is expected of her (as a participant in the script).

d. Delilah knows how to pick locks of such-and-such type and such-and-such make.

Not all of these scripts are generic. In fact, (a) and (b) are not generic at all, since they represent particular things. On the other hand, (c) and (d) are more general, and so more abstract. Also, (b) and (c) are likely to be heavily based on propositional or linguistic information; (a) and (d) are not. The point is that scripts will vary in the degree to which they may be considered “concrete” or “abstract,” and the information they carry will vary in both the mode of representation and complexity.

---


9 Regarding my use of “inference,” see point (xii) in §1.4.
Chapter IV  
Scripts I: Foundation and Structure

i. Long-Term Memory (LTM)

Most of the work of Schank and Abelson revolved around episodic scripts, but (as indicated above) I wish to develop their basic principles to cover more than just events and situations. In all, we shall examine four kinds of scripts: episodic, instrumental, personal, and definitional. These scripts are associated with at least three different long-term memory systems: episodic, procedural, and semantic.

The ideas developed in *Scripts, Plans, Goals, and Understanding* in relation to the theoretical entities of both CD and episodic scripts were meant to form the basis of a kind of LTM called episodic memory. According to Endel Tulving,

episodic memory receives and stores information about temporally dated episodes or events, and temporally-spatial relations among these events. A perceptual event can be stored in the episodic system solely in terms of its perceptual properties or attributes, and it is always stored in terms of its autobiographical reference to the already existing contents of the episodic store. (1972:385-6)

Schank and Abelson believe that most human knowledge is organized episodically. They give the following example,

If we ask a man, 'Who was your girlfriend in 1968?' and ask him to report his strategy for the answer, his reply is roughly: 'First, I thought about where I was and what I was doing in 1968. Then I remembered who I used to go out with then.' In other words, it really isn't possible to answer such a question by a direct look-up. Lists of 'past girlfriends' do not exist in memory. Such a list must be constructed. The process by which that list is constructed is a search through episodes organized around times and locations in memory. (1977:19)

The following claims are examples of the sort of information stored in episodic memory: "I remember hearing a siren a short while ago, followed by a flashing red-blue light;" "Last summer, while vacationing in Honolulu, I met an eccentric native who made a living selling balloons;" "I remember that I have an appointment with my dentist early next week;" and "One of the words I am sure I saw in the first list I studied was CACOPHONY."

Episodic memory is typically contrasted with semantic memory. Semantic memory—associated primarily with definitional scripts—is the memory necessary for the use of language.

It is a mental thesaurus, organized knowledge a person possesses about words and other verbal symbols, their meaning and referents, about relations among them, and about rules, formulas, and algorithms for the manipulation of these symbols, concepts, and relations. Semantic memory does not register perceptible properties of inputs, but rather cognitive referents of input symbols. (Tulving 1972: 386)

Examples of the kind of information stored in semantic memory are claims like: "I remember that black gunpowder is made from potassium nitrate, sulphur, and charcoal;" "I know that summers are quite hot in Bangladesh;" "There are four months with names ending with the letter ‘y’;" "I know
that $5! = 120$;” and “I think that the association between the words CROQUET and RUGBY is stronger than the association between BIPED and ENCEPHALITIS.”

The distinction between episodic and semantic memory was first proposed by Tulving in an article unassumingly entitled, “Episodic and Semantic Memory.” Although the distinction has not been universally accepted mainly because it departs from the principle of parsimony—we shall see in the following chapter that the evidence for it outweighs the arguments of its critics.

The third LTM system—linked to instrumental and personal scripts—is procedural memory, or memory for skills. When people tie their shoes, put a backspin on a tennis ball, or drive a car, they are drawing on the resources of procedural memory. The procedural LTM store contains knowledge of how to do things, and is implicated is almost every physical thing we do, from the nearly automatic to the premeditated and deliberate. An example of the former is reading, which involves a set of complex but largely unconscious measures for decoding strings of letters and words; an example of the latter is a bank robbery, which involves multiple nestings of plans, strategies, tactics, goals, subgoals, and as many alternatives to each of these as you like. The machinery of procedural memory is especially interesting to study in the context of problem solving, where difficulties like functional fixedness (the tendency to use objects and concepts in the problem environment in only their customary or usual way) and negative set (the bias or tendency to solve problems in one particular way, even when a different approach might be more productive) are often manifested.

ii. Episodic Scripts

Causal chaining was introduced in §4.3.v as a useful way to account for any sequential flow of events. Since certain event sequences frequently unfold in the same order, it is reasonable to postulate that people have developed special systems to deal with such patterns. These patterns are what we shall call episodic scripts (scripts_E).

What distinguishes episodic scripts from other types is the fact that scripts_E are about ordinary events and situations. For instance, rather than list all the details of what happened at a particular dinner trip in a restaurant, memory simply lists a pointer or link to a standard “going to a restaurant” script_E, then “tags on” the details that deviated from the standard script_E in the memorial representation of the trip in question. (Figure 4.2 shows what a RESTAURANTE might be like.) Although this economy has the side effect of poor memory for some details, it is necessary for achieving any kind of understanding at all. The alternative is to have memory so saturated with information (both relevant and irrelevant) to the extent that search time becomes impractically long.
Chapter IV  Scripts I: Foundation and Structure

<table>
<thead>
<tr>
<th>Script:</th>
<th>RESTAURANTES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Track:</td>
<td>Coffee Shop</td>
</tr>
<tr>
<td>Props:</td>
<td>Tables, Menu</td>
</tr>
<tr>
<td></td>
<td>Check, Money</td>
</tr>
<tr>
<td></td>
<td>F-Food</td>
</tr>
</tbody>
</table>

| Roles: S-Customer, W-Waiter, C-Cook, M-Cashier, O-Owner |
| Entry Conditions: S is hungry, S has money |
| Results: S has less money, O has more money, S is not hungry, S is pleased (optional) |

Scene 1: Entering

- S PTRANS S into restaurant
- S ATTEND eyes to table
- S MBUILD where to sit
- S PTRANS S to table
- S MOVE S to sitting position

Scene 2: Ordering

- (Menu on Table)
  - S PTRANS menu to S
  - S ATTEND menu (inst eyes)
- (Menu not on Table: W brings menu)
  - S MTRANS signal to W (or waits)
  - W PTRANS W to table
  - S MTRANS "need menu" to W (inst SPEAK)
  - W PTRANS W to menu
  - W PTRANS W to table
  - W ATRANS menu to S

- S MTRANS food list to CP(S)
- S MBUILD choice of F
- S MTRANS signal to W (or waits)
- W PTRANS W to table
- S MTRANS choice of F to W (inst SPEAK)
- W PTRANS W to C
- W MTRANS (ATRANS F) to C

  (Two Options)

  - C MTRANS "no F" to W (inst SPEAK)
  - W PTRANS W to S
  - W MTRANS "no F" to S (inst SPEAK)

  (Two Options)

Scene 3: Eating

- C DO (prepare F)
- C ATRANS F to W
- W PTRANS W to S
- W ATRANS F to S
- S INGEST F

  (Two Options)

Scene 4: Exiting

- S MBUILD leave restaurant
- S MTRANS to W (W ATRANS check to S)
- W ATRANS check to S
- S ATRANS tip to W (optional)
- S PTRANS S to M
- S ATRANS money to M
- S PTRANS S (ALOC out of restaurant)

Figure 4.2: A representation of what a generalized RESTAURANTES might look like.

Though the script may seem complicated, many details have been ignored. In fact, entire possible scenes have been left out—e.g., the "waiting to be seated by the hostess" scene. Also, some event sequences—such as paying the check and gratuity—may differ from one establishment to another. None of these niceties is all that important, however, just as long as the main elements in the script are present. The human intellect is malleable enough to accommodate whatever extra details or minor variations may be encountered during the course of normal experience. A few features of the script should be noted. First, some restaurants leave the menus on the tables, others don't. The beginning of the second scene accommodates each of these two contingencies. Second, after the customer has made an order, we find two options: her request is either satisfied (the meal is delivered), or it is not satisfied (the meal is not delivered). Here again there are two options: the customer may leave the restaurant, or she may consult the menu once more and modify her order. Third, we also find two options after the customer has finished eating her meal (end of scene 3): she may decide to pay the bill and leave, or she may order more items. Finally, although leaving a tip is usually a part of the dining-out experience, it is an optional practice, and is therefore left as such. (Adapted from Schank and Abelson 1977).
The need for two special mechanisms becomes immediately apparent. First, since we must be able to refer to frequent event sequences in a sketchy manner—otherwise risk boring our audience with tangential minutiae—we must be able to recognize that some script has actually been referenced. Second, we require a mechanism to recover the steps that have been left out of the causal chain. Consider the following sentences:

(10) Eva went to a restaurant. She ordered Beef Wellington from the waiter. She paid the bill and left.
(11) Eva got on a bus. She fell asleep. She woke up in Red Deer.
(12) Samson was the pitcher. He struck out sixteen players in a row. His team won the game.
(13) Samson went to Adam's birthday party. Adam opened his presents. Samson ate cake.

These stories are understandable because they make reference to common scripts (eating in a restaurant, riding a bus, watching or playing a baseball game, and participating in a birthday party). Much more than what is given in each story is understood by the audience. The listener must (and does) fill in the parts of each story that were omitted, and this is done by implicitly or explicitly referring to the invoked script. In contrast, the connectivity of

(14) Delilah went into a restaurant. She saw a waiter. She went home.
(15) Delilah was walking in the park. She asked the taxi driver about the Champagne bottle. She threw away the red mittens and skipped along.

seems to be somewhat lacking. Story (14) appears to reference a script but never quite gets there. Did Delilah eat or didn’t she? We can’t tell from the story, and so the pointer to the script cannot be safely applied. The events seem disconnected because of uncertainty about whether or not the “restaurant” script should actually be instantiated. Story (15) does not reference a script at all. We are unable to understand the events of that story because we have no specific knowledge that connects parks, taxi drivers, Champagne bottles, and red mittens (not to mention “skipping along”) in a coherent way.

Just what script is invoked depends on two things: tracks and headers. To illustrate,

(16) John went into a restaurant. He ordered a Big Mac. He paid for it then ate it while driving to work.

Now compare (16) to story (10). The restaurant in story (10) has a waiter who is able to serve patrons beef Wellington. This dish is not offered at just any restaurant. The image of a sophisticated eating

---

17 This is explained by Stalnaker as the making of pragmatic presuppositions: “To presuppose a proposition in the pragmatic sense is to take its truth for granted, and to assume that others involved in the context do the same ... Presuppositions are probably best viewed as complex dispositions which are manifested in linguistic behavior. One has presuppositions in virtue of the statements he makes, the questions he asks, the commands he issues. Presuppositions are propositions implicitly supposed before the relevant linguistic business is transacted ... Since the presuppositions play such a large part in determining what is going on in a linguistic situation, it is important that participants in a single context have the same set of presuppositions if misunderstanding is to be avoided. This is why presupposition involves not only taking the truth of something for granted, but also assuming that others do the same” (1970: 38-9). Stalnaker develops these ideas further in “Pragmatic Presuppositions,” “Indicative Conditionals,” “Assertion,” and “On the Representation of Context.” Grice (1975) refers to pragmatic presuppositions as conversational implicatures, which arise in consequence of a “cooperative principle” in speech exchanges. One of the tenets of this cooperative principle is that the partners in a conversation must not make their contributions more informative than is required.
place—with valet parking, cloth table covers, candlesticks, and perhaps a wine cellar—comes to mind. In contrast, the restaurant in (16) is of a much smaller calibre. It is the sort of establishment to which you can go and pick up your meal from a drive-by window. Story (16) is therefore understood by a particular variant or track in the restaurant script: the “fast-food” track; story (10) is understood by the “fancy-establishment” track. Other tracks may include the “coffee-shop” track (Figure 4.2), the “cafeteria” track, and the “snack-bar” track.

Whenever we encounter a story or situation like (16), elements of the story or situation trigger the appropriate script. As a consequence, all subsequent events in the story or real-world experience are interpreted with reference to the script that has been activated in memory. The elements that activate a script are called headers. In a general sense, a header is nothing more than a prime, a trigger that activates a related body of knowledge. Thus headers like “hungry” and “waiter” will prime the restaurant script, thereby providing access to the entire body of restaurant knowledge. In the case of story (10), the “Beef Wellington” header activates the “fancy-establishment” track of the general RESTAURANT script.

Notice that (16) requires somewhat more specific information than a general restaurant script, since in a typical restaurant situation one does not usually eat in a car. Here, the “Big Mac” header is instrumental in activating the relevant track in the overall RESTAURANT script. If one has specific knowledge of this track, one will not be surprised that John ate his lunch on the way to work. But if one has never heard of a Big Mac before, there will be some difficulty in understanding why John remained in his car.

Not all headers are of equal predictive strength. There are four main types. First we have the precognition header, so called because it references a script on a subconscious level. For instance, the word “hungry” in the sentence “Adam is hungry” is a precognition header for the RESTAURANT script. Adam’s being hungry may not render the audience readily prepared for the restaurant context; but nor will they be completely unprepared for it either, since there is a relationship (albeit an indirect one) between hunger and restaurants.

The multi-conceptualization header refers to the kind of primes that may instantiate more than one script, and thus involves a certain measure of ambiguity. One example is,

(17) Adam went to meet Eva the waitress. He had to wait half an hour for the end of her shift.

The RESTAURANT script is signalled by “waitress,” but the story may just as well be about a simple meeting between Adam and his girlfriend. On the other hand, Adam may have sat in the dining room and ordered a refreshment while waiting. These are alternatives between which we cannot easily decide, because “waitress” plays a double role here, and hence, the designation multi-conceptualization header.

We make stronger predictions based on the facilitation header. This type of header is commonly encountered in inputs that refer to two or more contexts, of which one could be interpreted as a means of accomplishing the other. In “Adam took the subway to the restaurant” both the “subway” and “restaurant” contexts are invoked, since subsequent inputs about either would not

---

18 Or the internal conceptualization header, as Schank and Abelson call it.

19 Schank and Abelson use the name instrumental header, I prefer using facilitation header because this avoids possible confusion with instrumental scripts (described below).
be unexpected. However, we also recognize that subways are a means of reaching locales, in which more important script goals may be achieved. Consequently, “subway” here acts as a facilitation header for the restaurant script. In turn, we understand that the RESTAURANT is in some sense a facilitator of the business context in the sentence “Adam went to a working lunch.”

The notion of the location of events leads to the fourth and most strongly predictive type of header, the locale header. Many episodes are known to have a “residence,” a place or building or institution where they characteristically take place. The symbol of such a place (if there be one) could then act as an indicator of the context in which to interpret and anticipate events. An example of the locale header is McDonald’s Golden Arches, which signal the “fast-food” track of the RESTAURANT to the public.

That script-based inferences are constantly made is evident from the following story:

(18) John was feeling very hungry as he entered the restaurant. He settled himself at a table and noticed that the waiter was nearby. Suddenly, he realized that he’d forgotten his reading glasses.

Virtually all readers (listeners, viewers) will understand John’s dilemma: he is unable to read the menu. How do we know this? When the restaurant script is primed, a whole set of frames—i.e., details about specific events or objects within the script—also become active. When a detail comes along (e.g., the waiter), it is simply stored in the appropriate frame, but if a detail is missing (e.g., the menu), it is “filled in” from the active script. Even though no menu was mentioned in (18), we understood John’s predicament because a menu is the most likely thing in a restaurant for which a reading glasses may be required.

The menu is what is called a default value for the frame—i.e., the common, typical value that occupies the frame. The experiencer takes for granted missing details in an episode for which he has a script by (non-consciously) following a simple rule,

General Rule for Assigning Frame Values: If a detail is missing, assume the normal default value as specified by the relevant script; if a detail is available, replace the default value with the detail.

This rule suggests that in a story like (10), all the frames between entering a restaurant and exiting it will be assumed, not just those mentioned in the description of the episode. So our understanding system will treat example (10) as if it had actually been:

(19) Eva entered a restaurant. She found a table and sat down. She read a menu.
    She ordered Beef Wellington from the waiter. Her meal arrived. She ate her meal.
    She asked for the check and received it. She paid the bill. She left the restaurant.

People do not normally assume too much, however. Our general RESTAURANT makes it fairly safe to suppose that the events of story (10) must have happened in the manner described in (19). But did Eva enjoy her dinner? Did she have dessert or coffee? Did she order wine? Was she

---


21 Marvin Minsky (1975) uses the word “frame” in a manner very similar to the way Schank and Abelson use the word “script.” The latter authors, in turn, use “frame” differently from the way that word is used by Minsky. Readers familiar with Minsky’s frame theory should keep in mind these different usages and not confuse them.
satisfied enough by the service to leave a large tip? We simply don’t know. The purpose of “filling-in” details is to aid comprehension, not to embellish or exaggerate. That is to say that the frames of a RESTAURANTes are meant to be general; they do not contain any information about Eva’s specific dinner experience.

To recap, people record in memory representational models of events they have experienced in the past (scriptsE), which are then invoked—or activated, or instantiated, or retrieved—when a new experience matches the stored model. This provides a framework or context within which the new experience is understood. In the case of eating in a restaurant, the relevant track (or variant) of your RESTAURANTes tells you the order the events will take, who the central players are, and what you and they are supposed to do. ScriptsE are activated by headers (or primes), of which there are at least four: precognition, multi-conceptualization, facilitation, and locale. Episodic scripts are also heavily involved in comprehension in more passive situations. As listeners, we are able to extrapolate missing details in a story by instantiating and implicitly referring to the relevant scriptE. We do so by noticing what frames (or details) in the active scriptE have been explicitly mentioned, and then following the general rule for assigning frame values.

One more point before we move on. I think it is reasonably clear that episodic scripts must reflect what Searle calls local background or local cultural practices (1983: 144). The background is a set of mental capacities that enable higher-order cognition to take place. The deep background includes capacities common to all humans in virtue of their biological makeup—such as walking, eating, perceiving, taking account of the solidity of things and the independent existence of objects and persons. In contrast, the local background comprises the idiosyncrasies associated with a particular culture at a particular time in history. It includes opening doors for ladies, drinking beer from bottles, and the stances we take toward such things as cars, refrigerators, money, cocktail parties, and restaurants.

Within both the deep and local backgrounds, Searle distinguishes those aspects that have to do with “how things are” from those that have to do with “how things are done”:

I can, for example, intend to peel an orange, but I cannot in that way intend to peel a rock or a car; and that is not because I have an unconscious belief, “you can peel an orange but you cannot peel a rock or a car” but rather because the ... stance I take toward oranges (how things are) allows for a completely different range of possibilities (how to do things) from that which I take toward rocks or cars. (1983:144)

Analogously, my scriptE of restaurants (how things are) allows for a completely different range of possibilities (how to do things) from my scriptE of churches, theatres, and high school reunions. Of course, I am not born with these distinctions already in place. They emerge from the episodes I experience, from my causal interaction with the environment around me, and the capacity of my brain to produce and store useful models.

iii. Instrumental Scripts

When Searle spoke of “how things are done,” he was referring to the attitudes we have toward the things that surround us, not the step-by-step actions carried out to accomplish specific tasks. The latter involve a somewhat specialized kind of long-term memory which psychologists call procedural memory. As mentioned, procedural memory consists of the kind of scripts—instrumental scripts (scripti)—that contain information about how to do things. Aside from the overall function of the script, the biggest difference between a scriptE and a scripti is the number of participants: episodic scripts usually call for multiple actors, instrumental scripts typically have just one.
Many of the tasks represented by scripts are performed non-consciously, but they do not always start out that way. In popular psychological jargon, we can say that the acquisition of some scripts may initially require conscious processing, but as time goes by and the relevant skills are perfected, the tasks involved will become examples of automatic processing.

Consider examples of motor control to clarify these remarks. When a baby begins to take her first steps, it requires intense concentration—conscious processing, in other words. Even with this conscious involvement, however, performance will not be flawless. If you call the baby’s name out, she’s likely to look toward you, lose her balance, and keel over onto the floor. The distraction of hearing her name took away attentional resources that were being devoted to the job of walking. With enough practice, however, the baby can now walk with little or no conscious effort involved and is not thrown off balance when attention shifts to another task. Conscious control may still be necessary under unusual circumstances, though, such as walking on the uneven surface of the backyard or climbing stairs. Mastering the coordination for the accelerator, clutch pedal, and gearshift in a manual transmission car is another excellent example—you eventually are able to pay attention to other things, like the traffic around you, because your foot and hand movements are now under automatic control. (Ashcraft 1994:70)

Instrumental scripts are susceptible to Searle’s distinction between deep and local backgrounds. For instance, some scripts are learned by socialization—e.g., how to pick a lock or use a word processor—and thus belong to the local background. Other scripts depend on more elementary skills, which are related to our basic physiological constitution and capacity (the deep background). Consider walking again. Although a baby may need to concentrate when taking her first steps, the ability to walk upright is normally considered to be a product of human biological evolution, common to all healthy members of the species, and not learned in the manner that, say, Euclidean geometry is learned. What we may call deep instrumental scripts are thus likely to be at least partially hardwired, and so more instinctual than the local variety.

The difference between deep and local scripts is consistent with Dretske’s discussion of goal-motivated actions in Explaining Behavior. Dretske maintains that an individual’s conscious goals will invariably depend on the execution of numerous subgoals. The term “subgoal” is not intended to refer only to physical events that yield explicit representations. In fact, we are not aware of most subgoal states that may be active at a given time. To be aware of a subgoal state, the state’s underlying physical event must yield a representation at a high enough level to render it explicit. It so happens, however, that many subgoals will be generated non-consciously, the result of implicitly represented physical activity. At very low levels—the neuronal level, for instance—a subgoal state can be as simple as a “decision” to fire or not to fire. We are neither aware of the presence of this subgoal state, nor of how, when, or even if it is discharged. Hence, signals from the brain cause our muscles to contract and relax in various patterns, thereby moving our bodies, without our awareness of the mechanisms which generate these motions. “The factors which are causally responsible for one thing’s causing another,” says Dretske,

need not be responsible ... for the way the effect is produced. A chief executive officer can get his employees to do certain tasks without knowing or caring how they do them. Indeed, as we all know, a hierarchical arrangement of executive responsibility is an efficient way to organize command structures. (1988a: 133)

But there are conscious subgoal states. Schank and Abelson propose to analyze these states by introducing the theoretical entity of the plan. “A plan is made up of general information about how actors achieve goals. [It] explains how a given state or event was prerequisite for, or derivative from, another state or event ... After an event initiates an MBUILD, that MBUILD is the reason for the next action. What is MBUILDed is a plan” (Schank and Abelson 1977: 70).
Consider a plan for satisfying one’s hunger. A crucial step in executing this plan is to bring oneself close to the item(s) intended for consumption. This step involves a change of state—in this case, the state of proximity to a specific target. This is an example of the kind of objectives which, because of the idea of change, are called *delta-goals*—or D-goals, for short. Table 4.1 lists the D-goal representing a change in proximity along with four others used in CD theory.

**Table 4.1: A list of the D-goals used by Schank and Abelson.**

<table>
<thead>
<tr>
<th>D-Goal</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>D-PROX</td>
<td>CHANGING PROXIMITY: a change in one’s location relative to a specified object or person.</td>
</tr>
<tr>
<td>D-CONT</td>
<td>GAIN CONTROL: to achieve possession or change physical control of some object from one holder to another.</td>
</tr>
<tr>
<td>D-KNOW</td>
<td>ACQUIRE KNOWLEDGE: to successfully obtain new information regarding an object, state, event, or phenomenon.</td>
</tr>
<tr>
<td>D-SOCCONT</td>
<td>CHANGE IN SOCIAL INVOLVEMENT: power or authority to do something; a change of state—of which inclusion and exclusion are two examples—in regards to some social position, event, or experience. This may involve such things as buying tickets for a performance, or requesting an invitation to participate in a party.</td>
</tr>
<tr>
<td>D-AGENCY</td>
<td>A CHANGE IN STATE BY PROXY: convincing a third party to pursue a goal on one's behalf.</td>
</tr>
</tbody>
</table>

Each of the five delta-goals is associated with a set of possible actions—called *planboxes*—that it “calls up” in order to be satisfied or fulfilled. There is a special set of planboxes—defined as a persuade package—which is involved in almost all D-goal states. The persuade package consists of ASK, INVOKE THEME, INFORM REASON, BARGAIN OBJECT, BARGAIN FAVOUR, and THREATEN. ASK is simply a request that some person(s) somehow aid your desired objective, and INFORM REASON is a statement providing some justification for this request. When resistance is encountered, INVOKE THEME may help to secure compliance by getting your opponent to sympathize with your plight through some kind of emotional appeal. BARGAIN OBJECT and BARGAIN FAVOUR are barter techniques. The first involves the exchange of an object for the desired goal; the second involves the exchange of a favour. Finally, when all else fails or is unlikely to work, THREATEN is sometimes a very effective means of persuasion. (Table 4.2 lists the planboxes for the five D-goals.) Here is an example.

---

22 Needless to say, no suggestion is being made here that these are the only goals associated with the idea of change or that delta-goals in general are exhaustive of the kind of objectives people might have. Rather, the simple list in Figure 4.1 is meant for illustrative purposes only. It is intended to show (using simple categories) the kind of structure a goal-laden instrumental script may have.
Samson wanted some of Adam’s apples. He said, “How about some of them for me?” (ASK). Adam refused. He reminded Adam what true friends they had been (INVOKE THEME), but Adam still refused. Samson said, “My In-laws are coming over for dinner and I need the apples to bake a pie” (INFORM REASON). Adam was almost moved, but not quite. Samson decided to casually put his big briefcase near the refrigerator and wait for his chance (STEAL). Meanwhile, he offered a bunch of bananas for the apples (BARGAIN OBJECT) and to cut firewood with his chainsaw (BARGAIN FAVOUR), but Adam still would not relent. Samson then implied that stinginess is incompatible with true friendship (THREATEN), and when that didn’t work, he became extremely frustrated and punched Adam in the nose (OVERPOWER).

Table 4.2: A list of the planboxes associated with each of the five D-goals.

<table>
<thead>
<tr>
<th>D-Goal</th>
<th>Planboxes</th>
</tr>
</thead>
<tbody>
<tr>
<td>D-PROX</td>
<td>Expressed in the form &quot;D-PROX(a, l),&quot; where a is the actor and l is the location. The planboxes for D-PROX are different from those of the persuade package. They are: RIDE ANIMAL, USE VEHICLE, USE PUBLIC TRANSPORTATION, and USE SELF.</td>
</tr>
<tr>
<td>D-CONT</td>
<td>Expressed in the form &quot;D-CONT(z),&quot; which means to gain control of physical object z. All the planboxes in the persuade package are suitable for D-CONT. Two more planboxes arise from the possibilities that one might gain control of the object without either the knowledge or permission of the owner (STEAL), or with his knowledge but without his permission (OVERPOWER).</td>
</tr>
<tr>
<td>D-KNOW</td>
<td>Expressed in the form &quot;D-KNOW(x),&quot; where x is always a fact or proposition. The actor for D-KNOW(x) is the person for whom the D-KNOW is a goal. The planboxes of D-KNOW are all of those in the persuade package, if x is a fact in the mind of the person from whom the information is solicited. If x is available in some other physical form (book, video or audio tape, etc.), then D-KNOW(x) involves two components: one must achieve D-CONT(x) followed by some decoding operation.</td>
</tr>
<tr>
<td>D-SOCCONT</td>
<td>The attainment of D-SOCCONT involves the same devices as those mentioned in the persuade package. Additionally, there are two planboxes roughly analogous to STEAL and OVERPOWER (in D-CONT)—namely, USURP and GO OVERHEAD.</td>
</tr>
<tr>
<td>D-AGENCY</td>
<td>This delta-goal is very flexible, since any goal or step in a plan can be pursued on one’s behalf by a third party. The planboxes for D-AGENCY are those of the persuade package.</td>
</tr>
</tbody>
</table>

As with the basic acts of CD, a question can be raised as to whether delta-goals and their planboxes constitute the right kind of primitives. With respect to the planboxes in the persuade package, Shank and Abelson claim that “when we ask people to write down all the ways to get control of an object someone else has, they inevitably give back a list which embodies almost all of these categories and no others. The distinction between BARGAIN OBJECT and BARGAIN FAVOUR is sometimes not made, but these kinds of bargains are clearly separate conceptually” (1977: 93). Be that as it may, we should deal with the raised concern in the same way we did for CD primitives: by
pointing out that, how well the hypothesized structures of scripts work, along with the plans to which they give rise, is something that can be tested using computer simulations. There is much to recommend this method, for it can reveal transparently and precisely how the various structural elements are implicated and utilized in a given task. We will look at one such program—called TAILSPIN—in the next chapter.

iv. Personal Scripts

The way we do things often betrays certain idiosyncrasies or personal habits. Some men shave using manual blades, others are inclined toward electrical razors; some wear their trousers before their socks, others insist that socks should be put on first. The instrumental scripts in such cases contain elements that are peculiar to the individuals whose scripts they are. These elements sometimes represent the passionate preferences of the individuals involved—preferences that may have nothing to do with the efficiency with which the tasks are performed. This suggests that there is more going on here than simply getting something done; it suggests the presence—or at least the influence—of what we shall call personal scripts (scripts<sub>P</sub>). Personal scripts are often manifested in the ritualized sequences of events in which individuals engage. Aside from a particular manner of shaving or getting dressed, these may include praying, preparing psychologically for a final exam, and getting ready to pitch a baseball.

Schank and Abelson contrast personal scripts with episodic scripts as follows:

Personal scripts do not behave in the stylized fashion of episodic scripts. All the participants in personal scripts are not necessarily aware of their participation. The seducee, say, or the victim of a swindle is often not aware until the very end of the enactment of the actor’s personal script of their participation in it. The personal script exists solely in the mind of its main actor. It consists of a sequence of possible actions that will lead to a desired goal... [The actor in a personal script] is participating in a sequence of events much like other sequences he has used many times before. He could teach his method to anyone who wanted to know it. There is very little planning involved because he has done this personal script repeatedly. (1977: 62)

23 I assume that ROBBERY is a personal script rather than an episodic script. I prefer the former designation because, unlike a generic script—e.g., RESTAURANT—there is no standard sequence of events which constitutes a robbery. How the theft is executed depends heavily on the actor’s goals and plans; the actor decides how the script is going to proceed, as it were, whereas a typical diner adheres much more closely to a pre-established set of norms and conventions.

24 But not impossible. Although D-CONT (money) is the main objective of Samson’s ROBBERY, there is more than one way to achieve this objective. For instance, if Samson’s STEAL planbox is not discharged satisfactorily (i.e., Samson cannot gain control of the money without raising anyone’s suspicions), he may have in mind OVERPOWER (inst gun) as a possible recourse.
The authors point out that clinical psychologists tend to be interested in very personal scripts that may be implicated in neuroses, especially when these scripts are activated inappropriately and create interference in the ongoing social behaviour of the individual. This is a case where the script concept may be useful to the demystification of abnormal behaviour, but we shall not pursue that matter here.

4.4 Definitional Scripts

*Definitional scripts* (scripts\(_D\)) do not refer to episodes, or tasks, or personal schemes and agendas. They pertain to the *nature of objects* (animate and inanimate), along with their *functions* and *relations*. As far as I know, scripts\(_D\) ordinarily do not constitute a part of script theory. But I think the extension is both natural and legitimate, and I want to say a little about it in this section.

It is very hard to imagine an episode taking place without any appearance of a physical object whatsoever. There are almost always players with specific roles involved, as well as props required for the proper execution of the scripts\(_E\). In the “coffee shop” track of the RESTAURANT\(_E\)s (Figure 4.2) we find tables, menus, checks, money, and various food items. We also have the customer, waiter, cook, cashier, and owner as the main characters. If children encounter episodes of this type at an early age, then they will also encounter whatever objects typically are associated with these scripts\(_E\) early in their cognitive career. Consequently, children will learn about the dynamics of the scripts\(_E\) at just about the same time they learn about the objects contained therein.\(^{25}\)

Consider a child who is taken to a restaurant for the first time. The events she experiences or witnesses there—being placed at a table, looking at a menu, ordering food, paying the check, etc.—will form the basis of an initial RESTAURANT\(_E\). During the trip the child is introduced to a variety of objects, including the waiter. Initially, the child may think that “waiter” is the exclusive title of the overweight, balding man who happened to serve her family on this occasion. Further trips to restaurants, however, will render her RESTAURANT\(_E\)s more sophisticated (e.g., she may come to learn that one must sometimes wait for a free table, that one cannot order food not listed on the menu, that it is customary to leave a tip at the end of the meal, etc.), while simultaneously rendering more elaborate her separate scripts\(_D\) for the objects associated with restaurants. Hence, a waiter will no longer be considered an overweight, balding man, but *any* individual who takes her order, brings the food, and who usually wears a uniform, speaks pleasantly, brings the bill, and is awarded with a gratuity at the end of the meal. In this way, the child’s development of a detailed script\(_D\) for waiter occurs alongside her development of a generic RESTAURANT\(_E\).

The above is an example of how a script\(_D\) of an *animate* object (waiter) might be generated. But definitional scripts also include inanimate objects, and the two should be treated separately.

### i. Inanimate Objects

Schank and Abelson believe that inanimate objects are defined by the scripts\(_E\) in which they appear. Thus a trip to the supermarket is stored in memory as a sequence of conceptualizations describing what happened on the trip. Some of these conceptualizations will be marked as salient; some will be forgotten altogether. Non-living objects fit in this view with a two-part definition. The first and primary part is a functional definition that attempts to generalize the salient roles the object

\(^{25}\) See §5.2.ii.
has played in past episodes. The complete functional definition of a given object lists all distinguishable occurrences of the object present in memory. The second part is a physical description of one particular member of the class that is being defined.

For a 'spoon,' ... the definition in memory lists the general usage for a spoon first (e.g., a thing that you PTRANS into mushy or liquid objects in order to PTRANS that object to your mouth so as to INGEST it). All interesting specific instances would also be stored (including, for example, 'The time I was camping and washed my spoon in the sand'). Last, we would have a physical description of a particular spoon (most likely the kind that you have at home). (1977:18-19)

The authors conjecture that for children, first experiences with an inanimate object tend to define it by establishing an initial script in which the item is placed. As the same object appears in more and more scripts, it might play different functional roles or display variant physical characteristics. This results in the child's concept of the object becoming more elaborate and sophisticated. Eventually, the object will take on a (definitional) script of its own, in something like the manner described for "spoon."

Notice that this is not inconsistent with the episodic-semantic distinction in long-term memory: to say that a definitional script of a particular object grew from the episodic experiences someone has had involving that object does not contradict that the relevant script may be represented in an LTM store other than episodic memory. As Tulving points out,

information stored in the semantic memory system represents objects—general and specific, living and dead, past and present, simple and complex—concepts, relations, quantities, events, facts, propositions, and so on, detached from autobiographical reference. If a person possesses some semantic memory information, he obviously must have learnt it, either directly or indirectly, at an earlier time, but he need not possess any mnemonic information about the episode of such learning in order to retain and to use semantic information. (1972:389; my emphasis)

Undoubtedly a great many things are defined in the context of episodes, particularly those that figure heavily in our everyday lives. But not all objects we know about have been introduced to us directly or by design in an episodic setting; nor is our acquaintance with every item restricted to physical features and functional properties. We have at least two good reasons to take seriously Tulving's allusion to the "indirect" learning of semantic information. To start, there are those things that fall outside the boundaries of "common" experience. We sometimes become aware of such things accidentally, perhaps by word of mouth or by reading an article, watching a documentary, or some kind of chance encounter. Secondly, many individuals—especially those we consider experts—can tell us, not only about what a given object looks like and what it does, but also about the object's microphysical composition and evolutionary history. So, if we consider the physico-functional knowledge gained of objects through their participatory role in episodic scripts direct, then it would appear that most of us acquire a considerable amount of information about things indirectly.

Keeping in mind the distinction between the structure of scripts and the manner in which information is gained will help to clarify these points. It is not unreasonable to suppose that at the earliest stages of development, humans begin to discover their environment in the context of

\[26\] Attribution theorists define an expert as one who, "by virtue of his modes of interaction with the environment, is capable of attaining a high information level in his own attributions and (unless other causal factors interfere) of making attributional statements presumed to have high validity. Trustworthiness implies the absence of irrelevant causal factors (personal motives, role demands) in the person's statements" (Kelley 1967:204).
personal episodes. But this does not mean that scriptsD are incapable of being informed in other ways. We are, after all, a species capable of forming complex social relations and communicating in verbal and written languages. So it should come as no surprise that our knowledge of things may, and often does, come from a multiplicity of sources. To say that the identity of the objects around us is represented in the human brain in the form of scriptsD does not commit us to a particular manner in which such knowledge comes to be possessed. And once this is understood, there should be no difficulty in seeing that conversations, magazine articles, documentaries, chance encounters, and the like, do not threaten the notion of scriptsD, but represent different ways of contributing to them. 27

ii. Animate Objects

Our scriptsD of animate objects will include both animals and humans, and will vary in sophistication depending on the age and mental capacity of the individual involved.

Animal ScriptsD Consider a child’s GIRAFFE$. It is likely to be somewhat superficial, consisting mainly of a physical description acquired in some non-direct way—e.g., a visual image from a magazine or television show. The description will be something like: “a large deer-like animal with long legs, a long neck, short horns, and dark brown spots all over its body.” Depending on the age of the child, the fact that giraffes are mammals living in Africa may also be included. A well-informed layperson may know yet more—e.g., that giraffes live mostly in east Africa, that they are not only the tallest mammals but also the biggest ruminants, that there are eight recognized races of giraffes, etc. A specialist is likely to have the most elaborate GIRAFFE$ of all, containing all sorts of “technical” information:

Very long neck with short, upstanding mane; high shoulders sloping steeply to hindquarters; long legs nearly equal in length. Male weigh 2420-4250 lbs; shoulder height 9-11 ft, top of horns up to 18 ft; females weigh 1540-2600 lbs; female shoulder height 2 ft shorter than males. Head tapers to point; long, prehensile tongue. Horns: solid bone, skin covered; a main pair in both sexes but female’s thin and tufted; male’s thick and bald on top, up to 5 inches long. A median horn and 4 or more smaller bumps in males ... Color brown to rich chestnut (old males darker, even black), dissected into intricate tapestry by patches and blotches of lighter hair; pattern unique in each giraffe. Scent glands: possible glands on eyelids, nose, lips; adult males have pungent smell. Teats: 4. 28

Experts who study giraffes in the field will have first-person experience of the animal. Such direct experience will inform the expert’s GIRAFFE$ much more deeply than, say, the knowledge a layperson might get from encountering a giraffe during the course of a ZOO$. But experts also study each other’s work; they read the research conducted by other specialists and evaluate their theories. In this way, much of the expert’s GIRAFFE$ will be constructed from what we have called “indirect” information, and, along with the technical information above, may contain details about the giraffe’s evolutionary lineage, ecological needs, daily activities, social and sexual rituals, offspring and maternal habits, and so on. Though direct experience in research has definite advantages, none of the expert’s knowledge suggests that it was acquired directly in an episodic context.

27 Similarly, Cullingford observes that a source of scriptal knowledge “is simply learning in school, from reading or from the media ... Even a young person who has never voted has some idea about political parties and candidates, and the event sequences involved in campaign swings, primaries, nominating conventions and the like. An adult who has never been to a fancy French restaurant may still know about the typical mode of dress, the haughtiness of the maitre-d, and the ritual of the wine” (1976: 8-9).

Generally, definitional scripts of animate objects will be more complicated than scripts of inanimate objects. There are at least three reasons for this. First, animate objects grow (and so change) in a manner that non-living objects do not. A giraffe calf will behave much differently from an adult animal, and will also have peculiar physical characteristics. Secondly, animate objects are much more interactive than inanimate objects. They appear to have "personalities" which sometimes result in behavioural idiosyncrasies; they react to stimuli in recognizable ways: by displaying signs of fear, aggression, curiosity, joviality, melancholy, and affection. As a consequence, our \textbf{GIRAFFE} is likely to be closer to our scripts of a person than, say, a \textbf{BRICK}.

The above brings to mind Dennett's \textit{Intentional stance}. Dennett argued that we deliberately view humans, animals, and sometimes machines as functioning in terms of a belief-desire system; and we do so because taking up the Intentional stance is a useful way of predicting behaviour. I doubt that there is anything deliberate about taking up the Intentional stance, but, without endorsing the eliminativist dimensions of Dennett's views, I would suggest that our adopting the Intentional stance for non-human objects may be explained as the \textit{non-conscious proclivity toward developing, for the objects in question, scripts that share many features with the definitional scripts we generate for people}. And this proclivity may be stimulated by (1) the measure of interactivity exhibited by a given object, and (2) the extent to which we are able to \textit{interpret} (correctly or incorrectly) the behaviour of the object as demonstrative of an independent will. Animals certainly fit these criteria.

The Intentional stance is especially pronounced in the case of those animals with which we share our lives. The relationships we have with our pets rank among the most intimate, and the definitional scripts representative of our knowledge of them are multifaceted. The reason for this suggests a third factor that may account for the difference between our scripts of living object and our scripts of non-living objects: the former tend to provoke \textit{emotional responses} from us more often (and usually with a higher degree of intensity) than the latter.\textsuperscript{29} Our pets really do seem to have minds of their own, to the extent that we sometimes treat them like infants or young children. That is not to say the emotional bonds are forged \textit{exclusively} with animate objects. Generally, a person will attach "meaning" to anything that may represent the ties (s)he has with special individuals. So, for instance, there is a decided emotional dimension associated with the pocket-watch my grandfather gave me for my high school graduation, and that dimension will occupy a place in \textbf{MY-POCKET-WATCH}. The difference, however, is that the feelings I have toward the pocket-watch have more to do with my grandfather than with the physical watch, whereas the emotions I have toward my pet canary are directed toward the canary itself. Unlike the watch, the bird reacts to and interacts with me, making it a much easier target of anthropomorphism.

\textbf{Human Scripts}\textsuperscript{3} This is where things get more interesting, and a lot more complicated. For there is a tremendous variety in the degree to which we are acquainted with others, and this variety is manifested in the intricacy of the definitional scripts we develop for the people around us. The first thing to note is the simple truth that we have absolutely no knowledge of most humans on this planet. I know that there are about one-and-a-quarter billion people living in China, but I haven't the faintest clue who they are or what they look like. Well, \textit{almost} no clue: I do hold a somewhat flimsy stereotypical idea of what a person from that region of the world might look like, and a still

\textsuperscript{29} Psychologists now speak of \textit{emotional memory} as a specialized mechanism in the LTM system. Recent research suggests that emotional memories are qualitatively distinct from more neutral memories and rely on different neural mechanisms. It has been found that the amygdala (a small region in the brain situated just below the prefrontal cortex) plays a crucial role in emotional processing, and damage or temporary impairment to the amygdala, through accident or drugs, leads to an impaired ability to recall emotionally salient events and to associate emotions with specific stimuli, behaviour, and objects (Westen 1999: 270-1).
flimsier idea of how (s)he might behave based on what I know of the culture, but I certainly cannot boast of anything more substantial.

My CHINESE-PERSON is therefore very superficial. It is not fine-grained enough to distinguish a person from China from, say, a person from Malaysia. But this can change with experience. The more I learn about the two races, the more sophisticated my concepts of them will become. Eventually, I may start to recognize differences in the relevant individuals based on physical features alone. One supposes that this is the case for those who grew up in either country. As it stands, however, my CHINESE-PERSON provides only a very general context for interaction. It enables me to make tentative predictions—regarding physical features, behavioural proclivities, tastes, values, etc.—which provide a rough guide for social situations.

There is nothing inherently prejudicial in forming such scripts. It just so happens that our minds are constantly looking for patterns, and when patterns are perceived—whether they be real or imagined—we cannot help but "catalogue" them. On the other hand, it is probably true that such constructs are at the root of many racist mind-sets. Problems arise when we fail to recognize the frailty of our stereotypes, when we try to apply them inflexibly to each single member of a particular group, thus implicitly equating all of them. When orthodox expectations are frustrated, our preconceptions and attitudes will dictate how we react: some might view individual uniqueness as atypical of the relevant race; others might look upon such uniqueness as an opportunity for augmenting and expanding the associated definitional script. Drew Weston (1999: 790) expresses both the necessity and danger of social schemas:

Schemas are essential for social cognition. Without them, people would walk into every new situation without knowing how to behave or how others are likely to act. Schematic processing can go awry, however, when schemas are so rigidly or automatically applied that they preclude the processing of new information. This often occurs with stereotypes, characteristics attributed to people based on their membership in specific groups. Stereotypes are often overgeneralized, inaccurate, and resistant to new information. Like other schemas and attitudes, however, they save cognitive "energy"; that is, they simplify experience and allow individuals to categorize others quickly and effortlessly.

In much the same way that biologists need not go to Kenya and observe giraffes in the field in order to form definitional scripts of the animal, we need not actually have direct experience with people in order to form scripts of them. This is obviously true in the case of famous historical figures. To take an example, I was first introduced to the writing of Charles Dickens at the age of fourteen, when, at the time, a simplified version of A Tale of Two Cities was a part of the English Literature curriculum at the school I attended. I immediately became impressed by the book and interested in its author. Over the next few years, I managed to read most of Dickens' original work, but somehow that was not enough. I wanted to know more about the man, and so decided to read what was then considered the two most reliable biographies: John Forster's The Life of Charles Dickens and Edgar Johnson's Charles Dickens: His Tragedy and Triumph. I learned a great deal about Dickens' childhood, education, love interests, the circumstances surrounding the publication of his novels, his social demeanour (which many considered to be aloof and vulgar), his trips abroad (including those to the United States and Canada), his exhausting work habits (which eventually contributed to his untimely death). I also learned about Dickens' associations, including his close friendship with Wilkie Collins, and was thus introduced to Collins' own tragic life. All this information—which includes Dickens' physical appearance as represented in portraits—comes together to form my DICKENS.

My introduction to Wilkie Collins in the course of studying Dickens' biography gives us a glimpse of how different scripts may be cognitively connected. Because Dickens and Collins were
friends, both renowned authors living at roughly the same time and location, their being cognitively associated may not seem strange. Other associations are far less natural, however. There need not be any rhyme or reason to how we come to link people or things together. Much depends on the context in which the connection is made. Figure 4.3 relates a rather odd personal example.

Because my knowledge of Dickens was obtained indirectly, my \textit{DICKENs} will be accurate only to the extent that the sources from which I have been informed imparted true information. But there is no good reason to believe that the account we currently have of Dickens' life is grossly inaccurate. To be sure, it is incomplete and probably contains some theorizing and guess work. Still, most respected biographies are based on reliable information: public records, testimonies and diaries, and Dickens' own extensive correspondence. So even though I have never met Charles Dickens, I am fairly certain that my \textit{DICKENs} gives me a reasonably accurate, if somewhat spotty, picture of the kind of man he was.

But not all scripts are so faithful. It is a remarkable fact of present-day society that some people have the power to command and manipulate the definitional scripts others form of them. The icons of popular culture—actors, singers and musician, sports personalities, politicians, and sometimes executives and industry leaders—all have armies of publicists and image consultants to ensure that the manner in which they are portrayed to the general public is well controlled. The picture of the incumbent politician kissing a baby during election time is a good example of this kind of manipulation. It is perfectly amazing that such devices still work!

In fact, sometimes they work a little too well. Most of us realize that the scripts we develop for certain people have little in common with how these people actually are. But we are still surprised when information contrary to what we "know" circulates. When it became known that River Phoenix died in 1993 of cardiac arrest due to a drug binge lasting nearly twelve hours, many people were shocked. Phoenix had presented himself as a health-conscious environmentalist, who worked hard for the welfare of several animal species and rejected the consumption of meat in favour of a vegetarian diet. Apparently, few suspected that this might just be a well-managed facade to which the actor's private habits bore little resemblance.

It is hardly surprising that the polished images public figures present cause the admiration, perhaps even infatuation, of many fans. Things take an ominous turn when admiration turns to
obession. In many such cases, no sharp distinction is made between the public persona and the private personality. And when the public image is finally discovered to be less than authentic, violence sometimes ensues. Mark David Chapman’s murder of John Lennon, Robert Bardo’s murder of actress Rebecca Schaeffer, Yolanda Saldivar’s murder of Latin singer Selena, and Arthur Jackson’s nearly fatal attack on actress Theresa Saldana (which led to California’s first anti-stalking law) are just a few examples. In all these cases, the perpetrators reported profound disappointment—eventually turning to rage—toward their victims: John Lennon was allegedly a “fake,” living a lavish lifestyle that was contrary to his professed philosophy; Rebecca Schaeffer had apparently “lost her innocence;” while Selena and Theresa Saldana were becoming “corrupt” by fame and fortune.

Ironically, the scripts we develop of media icons—people with whom most of us have never had direct (or even sincere / indifferent) contact—are often richer and much more emotionally charged than the scripts we have of many people we do meet. Those we encounter within an episodic script—i.e., the various role players—generally will come to mean little to us beyond the context of the script in question: my family doctor and his receptionist, the redhead from whom I occasionally buy coffee, the vacuum salesman who dropped by my home a while ago, the elderly lady I see waiting for the bus every morning, and most of the people with whom I took college courses are all examples. The scope of our knowledge of such people will be confined to the roles they play in the scripts in which they appear, and these role-players will in turn help to organize the associated episodic scripts.

On the other hand, common encounters sometimes lead to more significant relationships. Suppose that one of the players in an episodic script wants to direct the situation into channels other than those defined by the script. Suppose that in a RESTAURANT the customer decides to ask the waitress for a date. In this case, his behaviour is likely to differ from the typical restaurant goer, and will probably be defined by a personal script concealing a scripti with various subgoal states: he may try to engage the waitress in small talk or solicit personal information in the hope of finding out if she is attached (D-KNOW); he may try to present himself in the best possible light, e.g., by boasting about his occupation or personal possessions (BARGAIN OBJECT); or he may ask a friend to put in a good word for him (D-AGENCY).

It is possible that the waitress may ignore all of these overtures. But if she responds positively, a new relationship between her and the customer will have been instigated. Romantic attachment is one possible outcome, friendship is another. At least two factors distinguish close relations from passing acquaintance: the degree to which we are familiar with the other person and the intensity of the emotional bond we feel toward the person in question. We tend to know more about our intimate relations than other people. We know what they like and don’t like, their habits and quirks, their strengths and liabilities. And notice that this is precisely the kind of information we look for in those with whom we want closeness. On this view it takes (or should take) some time for a couple to discover whether or not they are compatible; and the longevity of their association will depend on the degree to which their respective scripts are similar and the ease with which each is able to make compromises when there are clashes. Contrary to popular opinion, opposites are less likely to attract.

30 Here is where intuition is likely to betray common-sense. Russell offers the following words of wisdom and warning: “Apart from self-knowledge, one of the most notable examples of intuition is the knowledge people believe they possess of those with whom they are in love: the wall between different personalities seems to become transparent, and people think they see into another soul as into their own. Yet deception in such cases is constantly practised with success; and even where there is no intentional deception, experience gradually proves, as a rule, that the supposed insight was illusory, and that the slower more groping methods of the intellect are in the long run more reliable” (1917:19).
In the case of friends and romantic interests, affection usually develops after we find out something about the individual, sometimes as little as physical attributes. But in the case of family, affection seems to be in place regardless of the characters of the people involved. Of course, the more we get along with certain family members, the more we become fond of them; that is why we have favourite aunts and cousins. But, generally speaking, our family relations will command some kind of emotional bond from us simply by virtue of being family relations, and this usually translates into an advantage over mere friendships. Some speculate that there is a genetic explanation for this. Most of the reasons offered for this belief involve cases of altruism. Others have argued that people usually have selfish reasons to behave altruistically—in order to relieve empathetic distress, the negative feelings (pity, compassion, concern, fear) aroused through empathy with another person in anguish.33

Evolutionary psychologists have united these two points of view by redefining self-interest in terms of reproductive success. In the bipartite essay, "The Genetical Evolution of Social Behaviour," W. D. Hamilton expanded the basis of reproductive success to include relatives who become ancestors of offspring with shared genetic material. This new inclusive fitness theory predicts that natural selection will favour those characteristics that cause a person’s genes to be passed on, regardless of whether or not the person in question produces offspring. If inclusive fitness is sound, one would expect humans and other animals to care preferentially for themselves, their offspring, and their kin. By assisting a genetic relation in time of need, one would help his or her relative become an ancestor of offspring who would share some of the helper’s genes. This is adaptive from an evolutionary point of view, since it is not the survival of the organism that is critical for natural selection; it is the survival of the genetic material. In short, despite the catchphrase “survival of the fittest,” inclusive fitness tells us that the ultimate criterion which determines whether a gene is spread is whether a given behaviour benefits the gene itself, not whether it is of benefit to the bearer of the gene. This explains the strength of familial bonds, and why we might feel closer to our kin than anyone else.

Is There a “Me” Script? Our discussion seems to be heading toward what may be the most intriguing representation of all: some kind of script defining selfhood. To be honest, I am not sure whether the basic apparatus of script theory can handle a generalized SELF, primarily because such a structure is liable to have a very different organization from the ones described so far. Dennett once defined personal identity as “the centre of narrative gravity” (1991: 418); I would amend this formulation to “the centre of episodic gravity.” “I” is that thing which binds our phenomenal universe together, not something that constitutes part of our experience. It is that thing to which events happen, not something that happens. Perhaps this is why we know ourselves much better

---

31 There are always exceptions, especially in cases involving abuse, cruelty, or neglect.


33 The claim that people sometimes render aid to others with an ultimate goal of benefiting the individuals to whom the service is directed is called the empathy-altruism hypothesis; the claim that people offer help in order to reap some personal benefit—e.g., acquire financial or social recognition, escape personal shame, avoid third-party chastisement, reduce empathetic distress, etc.—is called the egoism hypothesis. Batson and Shaw (1991) discuss both of these competing theories insightfully in an article entitled, “Evidence for Altruism: Toward a Pluralism of Prosocial Motives.”

34 Tulving (1972: 389) makes the following similar point: “A person’s episodic memories are located in and refer to his own personal past. Most, if not all, episodic memory claims a person makes can be translated into the form: ‘I did such and such, in such and such a place, at such and such a time.’ Thus, an integral part of the representation of a remembered experience in episodic memory is its reference to the rememberer’s knowledge of his personal identity.”
than we know anyone else, and why our self-knowledge is seldom the result of deliberate exploratory effort (psychoanalysis notwithstanding).

Yet I do have a self-concept which grows, changes, and is more influenced by how others perceive me than I would care to admit. Statements like "I never thought myself capable of that" suggest that we sometimes learn about ourselves in the same way we learn about other people: by observing and then evaluating how we behave in certain situations. Furthermore, some of the basic information my self-knowledge encompasses—physical appearance, preferences in food, art, literature, and music, closest friends and favourite kin, memories of specific objects, occasions, and circumstances—do not appear to be beyond the handling capacity of script theory. So at least some aspects of personal identity are amenable to a scriptal organization.

Douglas Hofstadter regards selfhood as a product of a struggle between a kind of idealized image and the external evidence which inevitably points to something less than perfect.

One of the most severe of all problems of evidence interpretation is that of trying to interpret all the confusing signals from the outside as to who one is... The psychic mechanisms have to deal simultaneously with the individual's internal need for self-esteem and the constant flow of evidence from outside affecting the self-image. The result is that information flows in a complex swirl between different levels of the personality; as it goes round and round, parts of it get magnified, reduced, negated, or otherwise distorted, and then those parts in turn get further subjected to the same sort of swirl, over and over again—all of this in an attempt to reconcile what is, with what we wish were... The upshot is that the total picture of "who I am" is integrated in some enormously complex way inside the entire mental structure, and contains in each one of us a large number of unresolved, possibly unresolvable, inconsistencies. These undoubtedly provide much of the dynamic tension that is so much a part of being human. Out of this tension between the inside and outside notions of who we are come the drives towards various goals that make each of us unique. Thus, ironically, something which we all have in common—the fact of being self-reflecting conscious beings—leads to the rich diversity in the ways we have of internalizing evidence about all sorts of things, and in the end winds up being one of the major forces in creating distinct individuals. (Hofstadter 1979:695-6)

Hofstadter’s "upshot"—that personal identity is so well integrated into one's entire mental framework that a separate SELF may be inextricable—is certainly a possibility, and one worthy of consideration. On the other hand, the idea of a deeply integrated selfhood does not necessitate the absence of a sufficiently precise SELF; it may, rather, be indicative of the enormous complexity of such a structure. The issues are obviously difficult, and demand a much closer analysis than I can provide here. So I am content to register my uncertainty as to whether the scriptal approach can shed light on the problem of personal identity and simply leave the matter at that.

4.5 The Basic Structure of Scripts

We have already discussed the structure of episodic scripts in §4.3.ii, where elements such as headers, tracks, frames, and default-values were identified and explained. These elements are also present in instrumental, personal, and definitional scripts.

---

35 No behaviourism intended!
### i. Headers and Primes

First, all scripts, regardless of type, have headers. Headers identify what a particular script is about. The headers I have used in this chapter were chosen for maximum clarity. But it is important to realize that other headers could have been selected, and that one and the same script can reasonably go by many different names.

We have made extensive use of the RESTAURANTES; others episodic scripts are CHURCH-GOINGES, DOG-WALKINGES, and DOCTOR-VISITINGES. Examples of headers for instrumental scripts include JObEs, GETTING-DRESSEDEs, MAKING-A-TELEPHONE-CALLEs, ROLLING-A-CIGARETTEEs, and CHANGING-THE-TIREEs. For personal scripts we have ROMANCEREs, ROBBERYEs, and SPYEs. The headers of definitional scripts are likely to be proper names where people are concerned (CHARLES-DICKENSes, WILKIE-COLLINSees), and natural-kind terms for many animals and plants (GIRAFFEEs, DOGEs, ELMe, ROSEEs). But we also have headers of particular objects and non-natural kinds (CHAIREs, MY-POCKET-WATCHEs, ATOMIC-BOMBes, and RESTAURANTEs).

A script is primed when some element displays one or more features to which the frames of the script are sensitive. This may happen in various ways: by words, visual images, sounds, smells, textures, signs and symbols. For a waiter script, primes might include a uniform, an apron, one or another functional role, or utterances like, “Can I take your order, please?” The header of a particular script is certainly enough to prime it, but whether or not the script becomes active depends on the context of the situation. For instance, a certain scene in a movie might stimulate Adam’s ROMANCERes, but that does not mean that Adam will immediately start to behave like a lover. He may wait until Eva is around, perhaps “taking notes” from the movie scene in the meantime.

All scripts—but scripts in particular—are primeable by descriptions. Hence, the portrait in Figure 4.3 primes my NAPOLEON-BONAPARTEes, but so does “the French emperor who was defeated by the Duke of Wellington at Waterloo in 1815.” Of course, if I did not know that Napoléon was the French emperor who was defeated by Wellington at Waterloo in 1815, my NAPOLEON-BONAPARTEes is not likely to be primed by this description. More on this in chapter VI.

### ii. Tracks

Tracks were described as variants on a particular script. We are already familiar with the tracks of the RESTAURANTEs. Now consider a person’s RESTAURANTEs, which amounts to that person’s concept of restaurant. There is no question that concepts can vary from person to person. Generally, an individual’s RESTAURANTEs will include tracks relating to fast-food establishments, four-star restaurants, coffee shops, cafeterias, and so on: these are the different kinds of restaurants with which most of us are familiar. We must also keep in mind that a scriptes is normally created from a particular role’s point of view. A customer sees a restaurant one way, an owner sees it another. So different perspectives will also make for different RESTAURANTEs tracks.

---

36 Note the difference between RESTAURANTEs and RESTAURANTEs.

37 Speaking of movies, it is interesting to observe how some directors increase suspense by keeping the audience guessing about the personal scripts and the nature of goal and subgoal states of the main character(s). This technique is most often used in murder mysteries and thrillers, where the viewers may expect many “twists” in the plot.
Tracks for human subjects also exist. These often consist in the various roles or functions people perform. For instance, though their basic duties are the same, one would expect the defining features of a waiter working at a sidewalk café to be different from those of a waiter working in a four-star restaurant: the latter, one expects, will be more formally attired, better groomed, more sophisticated in his mannerisms, and so forth. So there is a parallel between the different tracks of a RESTAURANTs and the different tracks of the role-players (e.g., WAITERs) that we might find in a restaurant setting. We also have different tracks for those with whom we form close relationships. My best friend is a teacher, a father, a husband, a brother, a son, and a member of a certain race. Over the years, I have had the opportunity to observe and learn about him in all of these roles, from which my “entire view” of him has surely benefited; and I can, without much difficulty, think of him in one of these ways rather than the other.

It is interesting to consider whether one might have different tracks of his or her own self-concept. The idea does not strike me as being strange. We frequently wonder what things would have been like if we were different in some way. Thomas Nagel (1974) wondered what it is like to be a bat!38 The fact that such reflections may be “philosophically problematic” is beside the point. In any case, most of us consider far less radical changes. I sometimes think about what I would have been like if I had studied musical theory instead of philosophy, if I were raised in a different country, if I had no brother, no sister, two brothers and two sisters instead of one of each. While these reflections may not show that we actually have different tracks of our self-concept, it does show that different possibilities of selfhood are sometimes contemplated, and probably more often than we consciously realize. Our ability to consider alternatives in this way is probably the source of both personal ambition and personal regret.

Lastly, we come to the tracks of personal and instrumental scripts. For both types of scripts, alternatives are often consciously adopted—e.g., “If our first plan to rob the bank turns sour, go to plan B,” or “If she does not react positively to your ‘take charge’ approach, give her the ‘sensitive’ routine.” The planboxes of D-goal states are, therefore, obvious candidates for variations in these scripts. But there are differences of which we must always be unaware. This is because the way we do things consists of myriad distinct physiological processes, involving an astronomical number of unique physical events over which we have absolutely no control. Still, though different instances of personal and instrumental scripts are almost certainly not identical “all the way down,” inter-token similarity can be found if we move high enough up the representational ladder: as mentioned, some men are fastidious about how they shave; for them, this process assumes the aspect of a step-by-step ritual from which they rarely deviate.

iii. Frames and Default-Values

The frames of a script are its complement of informational bits or details, which naturally vary from one type of script to another. The frames of an episodic script will contain information about the relevant episode—sequence of events, role of participants, function of props, etc.—and this provides appropriate anticipatory guidance to all the players involved. What makes scripts especially apt for predicting behaviour and events is the idea of default values: if there were not “stable” frames that are assumed until explicitly dislodged by contrary data, we would not get the power of forecast so necessary to our social nature.

38 See Hofstadter’s (1981) commentary on Nagel’s famous article.
The frames of *instrumental* scripts contain information about how things are done. Again, where physical movement is involved, frames will be extremely variable if the relevant physical sequences are considered at a sufficiently low level. But at higher levels—say, the level of decision-making—we find much more stability: there are particular ways things are accomplished, and we generate memorial representations in the form of scripts, not only of how each separate event of a given task is to be carried out, but also of its proper place in the event sequence and its relation to other events. It is helpful to keep Dretske’s analogy in mind here: “A chief executive officer can get his employees to do certain tasks without knowing or caring how they do them.”

Improvisation is obviously important, not only for scripts, but for personal scripts that contain instrumental components. Speaking generally, Abelson observes that

thoughtful processing can ... occur in script performance when obstacles or unusual variations occur. A script is more than just a linear list of inexorable events. In learning a script, one presumably learns variations in addition to constancies. (1981:723)

Fortunately, most of the impediments we encounter when doing things can be easily resolved—e.g., “the batteries in the remote control are dead” or “the milk jug is empty.” But although there is truth in Abelson’s observation, it is neither practical nor necessary to include all possible obstacles and contingency plans in memory representations of scripts and scripts. Surely my USING-THE-ELEVATOR is not to contain the frame “do not get off the elevator if it stops at the 5th floor,” all I have to know is the floor on which my apartment is located, and then I know both where to get off and where not to get off.

Of course, impediments to action are occasionally serious—e.g., “I want to mow the lawn but my neighbour’s ferocious dog is taking a nap on my front yard”—at which times we often devote some thought on how to resolve them, probably using a variety of different scripts in the process. Timing and quality of execution become critical when scripts representing “ulterior motives” are at stake. The characteristic we call *resourcefulness* is perhaps a function of the extent to which we are successful at creatively deploying various kinds of scripts to resolve barriers to action or advance personal goals.

As far as the frames of *definitional* scripts are concerned, these will be different for animate and inanimate objects. For inanimate objects they will normally consist in information pertaining to physical appearance and function; for animate objects they will also comprise behavioural features and, in many cases, involve emotional memory. The nature of the behavioural information contained in a script, and the extent to which emotional mechanisms may be involved, will depend on the definitional object itself. The average person’s WAITER, for instance, contains frames concerned with no more than appearance, function, and *role-related* attitudinal dispositions. In contrast, one’s definitional script of his or her grandma will comprise all these elements plus representations of the numerous salient memories of past interactions and all the emotions such memories invoke.

Once a particular script has been activated by a cue, the whole set of frames associated with that script is also activated. Hence, once we decide who the waiter is upon entering a restaurant, we then know that he is the one to whom all the elements which constitute the definitional script of waiter (or, better yet, the elements which constitute a particular *track* of that script) apply. Even when a specific element is not immediately present (e.g., the waiter has not yet taken my order), it is simply anticipated from the active script, if he is the waiter, then he will take my order. In all cases involving scripts, the rule for “filling in” missing information is the same: if an element is not
perceived, assume the normal default value of that element as prescribed by the relevant script. This is script theory's way of accounting for human inference-making.

iv. Scripts, Habits, and Roles

I want to close by distinguishing the script construct from two other theoretical entities in psychology. One of these is habit, a notion made popular by behaviourism. It may seem that scripts slyly sneak behaviourism in through psychology's cognitive door; and in a sense, this is true. Scripts play a double role in psychology: operationally (as in actually going to a restaurant) and cognitively (as in understanding situations involving restaurants). So while scripts are implicated in behaviour, they are also involved in cognition; and the cognitive aspect is considerable. The difference between a script and a habit is that a script is a knowledge structure, not just a response program, and thus there is access to it symbolically as well as through direct experience.

The way symbolic factors can enter a script is that a modified understanding of a script can result in modified behaviour, and the connection is causal. Thus Abelson gives the example, "when the veterinarian advises giving our dog a heartworm pill before chow, I try to change my dog-feeding script by adding this new step" (1981: 722). While a behaviouristic account might still call this a "malleable habit," it seems more natural to analyze the performance in terms of an enabling knowledge structure, in terms of a script.

Scripts must also be distinguished from role theory. Again, there are crucial differences despite some degree of similarity. Role theory tends to emphasize the web of social and institutional expectations constraining social behaviour, whereas a script-based theory fixes on individual cognitive structures that may or may not be congruent with the performance expectations of others. Of course, one can refer to the "customer role" or the "waitress role" in a restaurant (as I have done above), but in the context of script theory, the reason for doing so has to do with human comprehension, not social politics. That is to say that scripts, when well understood, will have anchorage in the psychology of cognition and memory, an anchorage not characteristic of role theory, which has its origins in sociology.

I have deliberately stayed away from Intentionality in this chapter. My purpose has been to lay the groundwork necessary for treating the philosophical issues raised in the first half of the dissertation. A little more work remains before we can take on these issues, however. The evidence for scripts, from both AI and psychology, is the topic of the next chapter. (See Figure 4.4 for a schematic summary of script theory.)

---

Figure 4.4. A schematic summary of script theory. Four kinds of structures are shown, of which scriptsD and scriptsi are the most complex. We form definitional script of both animate and inanimate objects. Some of the scriptal varieties within these two categories are represented, as are D-GOAL states and planboxes.

Conceptual Dependency Theory

Script Theory

Episodic  Instrumental  Personal  Definitional

Deep Background  Local Background  Deep Scriptsi  Local ScriptsD  Stylized Rituals  Hidden Agendas (Acts of Deception)

Animate Objects

Inanimate Objects

Animals

Plants

Persons

Pets  Non-pets

Cultural & Physical Stereotypes

Historical Figures  Public Personalities  Passing Acquaintances  Friends  Family Relations  "Me"

All Scripts Have the Same Basic Organization

Headers & Primes  Tracks  Frames  Default Values

Precognition  Multi-conceptualization Facilitation  Locale  Direct Info.  Indirect Info.

Plans  Delta goals

D-PROX  D-CONT  D-AGENCY  D-SOCCONT  D-KNOW(x)

RIDE ANIMAL    STEAL OVERPOWER  ASK
USE VEHICLE    INVOKE THEME  INFORM REASON
USE SELF  BARGAIN OBJECT
PUBLIC TRANS.  BARGAIN FAVOUR

x Is In The Mind Of An Agent  x Is In Physical Form ω

Acquire (D-CONT) ω  Decode ω
 Scripts II: Evidence from AI and Psychology

It will not be possible to apply exactly the same teaching process to a machine as to a normal child. It will not, for instance, be provided with legs, so that it could not be asked to go out and fill the coal scuttle. Possibly it might not have eyes. But however well these deficiencies might be overcome by clever engineering, one could not send the creature to school without the other children making excessive fun of it.

Alan Turing, "Computing Machines and Intelligence," p. 456.

The business of this chapter is to present support for the theory of scripts as it applies to human cognitive processing. The thesis for which I have been arguing (and will continue to develop) is that some, but not all, human Intentional states involve scripts. Accordingly, what I have to say below may not apply to all aspects of cognition or to all types of Intentional states.

The evidence I have in mind falls under two categories: computer task modelling and psychological studies. It often happens that corroboration for a hypothesis comes from different quarters and must be regarded cumulatively in order to be properly appreciated. I suggest that this is one such instance. When we have finished, there will still be unanswered questions and points of contention, very few positions are universally transparent. Nevertheless, I shall rely upon the reader’s personal experiences and common-sense (insofar as I can do so safely) to fill in the gaps and to lend assistance when my powers of description prove inadequate. May the partnership prove fruitful!

5.1 Computer Task Modelling

Aside from their interest in how knowledge is structured in memory, Schank and Abelson were involved in building computer software that can deal interactively with the public. These two pursuits are not unrelated, since one possible test for the adequacy of a theory of human understanding is its viability as a basis for computer programs. While AI modelling may not constitute definitive proof for the operation of scripts in the human cognitive apparatus, it is nevertheless suggestive, and must be considered accordingly.

Three computer models are described below: SAM, a story understander based on scripts; FRUMP, a newspaper story skimmer based on scripts; and TALESPIN, a storytelling program that uses scripts, plans, and goals.
i. SAM

SAM (Script Applier Mechanism) is a program that was developed at Yale to deal with stories that rely heavily on scripted sequences of events. SAM creates a linked causal chain of conceptualizations that represent the events described in the input text. This is done in two steps: the program first parses the story into conceptualizations using Riesbeck's (1975) Analyzer; these conceptualizations are then fed into a subroutine that looks for script applicability. When a script is found, it is used by a script Applier to make inferences about events that must have occurred between the episodes explicitly mentioned. Aside from being the first step in inference generation, the analysis of conceptualizations into causal chains also enables SAM to produce paraphrases of the input text. These paraphrases are normally longer than the original input, because SAM retains the inferences produced by the script Applier and incorporates them into its output. Abridged summaries may also be created using mechanisms that rely on measures of the relative importance of the events within the script. Finally, a sophisticated subroutine developed by Schank and Abelson allows SAM to answer questions about the story by examining the interconnections between the obtained conceptualizations and the inferences. These various components are described in detail below, but first let us look at a few illustrations (Schank and Abelson 1977: 178).

Input: John went to a restaurant. He sat down. He got mad. He left.

Output: JOHN WAS HUNGRY. HE DECIDED TO GO TO A RESTAURANT. HE WENT TO ONE. HE SAT DOWN IN A CHAIR. A WAITER DID NOT GO TO THE TABLE. JOHN BECAME UPSET. HE DECIDED HE WAS GOING TO LEAVE THE RESTAURANT. HE LEFT IT.

The first sentence in the input indicates to SAM that the RESTAURANTES is the applicable script in this case. The system can now "fill in" missing information. For instance, the RESTAURANTES contains the information that most people go to restaurants in order to eat, and so hunger must be a motivating factor for John; SAM thus infers that "John was hungry." The program also "knows" that before anyone can do x, (s)he must decide to do x. This general rule, along with SAM's causal chaining analysis, reveal two missing MBUILDs in the story: one representing John's decision to go to a restaurant, and one representing his decision to leave it. So the program adds these intermediate events in its summary. Moreover, stored in SAM's RESTAURANTES is the information that failure to get service can cause adverse reactions in customers—e.g., anger, leaving a small (or no) tip, complaining to management, etc. Since the input indicates that John became angry at just about the point where the active RESTAURANTES predicts the waiter to make an appearance, SAM infers that John

---

1 Cullingford (1976: 18-25) provides a good overview of SAM's operations.

2 Riesbeck gives a developmental history of the Analyzer, including a comparison with earlier systems that used similar components. He explains that the distinction between SAM's Analyzer and these other versions stems from a difference in philosophies: "The goal of the development of the Analyzer has not been one of building an immediately practical understander of the English language. Nor has it been an experiment to see how much can be done with a small, restricted set of functions. It has been a search for mechanisms that seem consistent with what people do when they understand language" (1975: 97-8). And a little later he observes: "Most other approaches, computational and linguistic and even psychological, have been concerned with what could be obtained using just syntax, until it becomes necessary to add semantics. The approach here is the exact opposite, to see what can be done from the conceptual side and only include syntactic aspects when they seem needed. The first form of the Analyzer didn’t even have word order" (1975: 111).

3 Of course, this inference may be wrong. John may have gone to the restaurant to get a beverage; or perhaps he went there to meet a friend. But although there is nothing definitive pointing to the fact that John was hungry, this is the most natural inference to make. As Schank observes, "there is no way of always being correct in natural language understanding simply because natural language is full of ambiguity. What we strive for is the right guesses for the right reasons" (1972: 593).
failed to get service. Here is another example that uses the question-answer subroutine (Schank and Abelson 1977: 178-9):

**Input:** John went to New York by bus. On the bus he talked to an old lady. When he left the bus, he thanked the driver. He took the subway to Leone's. On the subway his pocket was picked. He got off the train and entered Leone's. He had some lasagna. When the check came, he discovered he couldn't pay. The management told him he would have to wash dishes. When he left, he caught a bus to New Haven.

**Output:** JOHN WENT TO A BUS STOP. HE WAITED AT IT A FEW MINUTES. HE ENTERED A BUS. THE DRIVER GOT THE TICKET FROM JOHN. HE WENT TO A SEAT. HE SAT DOWN IN IT. WHILE JOHN WAS ON THE BUS AN OLD LADY AND JOHN TALKED. THE DRIVER TOOK JOHN TO NEW YORK. HE WENT TO THE DRIVER. WHEN GETTING OFF THE BUS JOHN THANKED HIM. JOHN GOT OFF IT.

HE ENTERED A STATION. HE PUT A TOKEN INTO THE TURNSTILE. HE WENT TO THE PLATFORM. HE WAITED AT IT A FEW MINUTES. HE ENTERED A SUBWAY CAR. A THIEF WENT TO JOHN. THE THIEF PICKED JOHN'S POCKET. HE WENT AWAY FROM JOHN. JOHN WENT TO THE SEAT. HE SAT DOWN IN IT. THE DRIVER TOOK JOHN TO LEONE'S. HE LEFT THE SUBWAY CAR. HE LEFT THE STATION.

HE ENTERED LEONE'S. HE LOOKED AROUND INSIDE IT. HE SAW HE COULD GO TO A TABLE. HE WENT TO IT. HE SAT DOWN IN THE SEAT. HE ORDERED SOME LASAGNA. THE WAITER INDICATED TO THE CHEF JOHN WOULD LIKE HIM TO PREPARE SOMETHING. THE CHEF PREPARED THE LASAGNA. THE WAITER GOT IT FROM THE CHEF. THE WAITER WENT TO THE TABLE. HE SERVED THE LASAGNA TO JOHN. HE ATE IT. HE BECAME FULL.

HE ASKED THE WAITER FOR THE CHECK. JOHN GOT IT FROM THE WAITER. JOHN READ THE CHECK. JOHN DISCOVERED HE WAS UNABLE TO PAY THE CHECK. HE INDICATED TO THE WAITER HE WAS UNABLE TO PAY THE CHECK. THE MANAGEMENT TOLD JOHN HE WOULD HAVE TO WASH DISHES. HE ENTERED THE KITCHEN. HE WASHED DISHES. HE LEFT LEONE'S.

HE WENT TO THE BUS STOP. HE WAITED AT IT A FEW MINUTES. HE ENTERED THE BUS. THE DRIVER GOT THE TICKET FROM JOHN. HE WENT TO THE SEAT. HE SAT DOWN IN IT. THE DRIVER TOOK JOHN TO NEW HAVEN. HE GOT OFF THE BUS.

Many of SAM's inferences can easily be identified in this long-winded paraphrase. Some of these seem unimportant—e.g., that John waited for the bus for a few minutes, that he put a token into the turnstile, that he read the check after he got it from the waiter—and would probably not have been mentioned at all by a human storyteller. Still, the fact the SAM is able to draw these inferences (all of which are reasonable, it should be added) is no menial accomplishment. The third paragraph clearly shows the extent to which the program's RESTAURANTEs is guiding its processing. Next, SAM answers some questions about the story (Lehnert 1977: 47-8).

**Input:** Why did John go to New York?
**Output:** BECAUSE JOHN WANTED TO GO TO LEONE'S.

**Input:** How did John get to Leone's?
**Output:** JOHN TOOK A BUS TO NEW YORK AND THEN HE TOOK THE SUBWAY TO LEONE'S.

---

4 Compare the events described in the last paraphrase to the generic RESTAURANTEs represented in Figure 4.2.
In order to understand how SAM is able to generate paraphrases and answer questions like these, we will need to take a closer look at its components.

The English Analyzer and the Script Applier Two closely related mechanisms in SAM are the English-to-Conceptual-Dependency Analyzer and the script Applier. The Analyzer takes the input text and extracts from it all the conceptual information conveyed by the linguistic elements. Later, different routines use the output of the Analyzer in CD and never deal with language again. Only the Analyzer grapples with problems of word meaning, inflections, ordering relationships, and other idiosyncrasies of linguistic expression.\(^5\) The Applier uses clues in the input to determine what script, if any, is to be used in processing the text.

Both the Analyzer and the Applier function by making predictions. The Analyzer uses the words in the input to access routines—called expectations—that predict what conceptual and linguistic forms are likely to occur later in the text. These expectations also specify what Conceptual Dependency structures should be built when the predicted forms are encountered. Combinatorial explosion is an issue, here, so the Analyzer must be able to make guided assumptions about word meaning in order to reduce the number of possibilities. By embedding expectation routines within

---

\(^5\) Riesbeck relates how the evolution of the Analyzer component reached an operational level sophisticated enough to handle some rather complex task domains. “The domain was the game Diplomacy. Diplomacy, although a board game, differs strongly from others like chess ... in that it depends heavily on interpersonal communication. Further, success in the game depends primarily on being able to influence, honestly and dishonestly, other people to do what you want, and at the same time judge how and why someone else is trying to use you. Hence there is a great deal of concern with human perception and communication of human behavior” (1975: 86). The Analyzer was thoroughly tested against Diplomacy before going into SAM.
CD forms, which are in turn embedded within larger script structures, the Analyzer uses general world knowledge and language-specific cues to determine the meaning of ambiguous expressions.

For instance, to understand that "the check came" means that the waiter brought the check requires knowledge of who does what, when, and to whom in restaurants. The Analyzer accesses this information by working closely with the script Applier. The partnership generates a situational context within which the meaning of words can be ascertained with reasonable efficiency. Generally, when a given script is activated by the Applier, the Analyzer uses that script to aid comprehension by accessing verb tenses that are situation-specific. Thus the analysis of "The waiter served ... " in the restaurant context never encounters the sense of "serve" that is appropriate to a tennis game or a military situation, unless, of course, the input forces it: "The waiter served in the army for five years." Situation-specific senses of nouns—e.g., "check" or "buck"—are handled similarly (Cullingford 1976: 20).

It is here, at the phase of text analysis, that SAM undertakes its most complex operations. Along with the Analyzer and the Applier, the system makes use of a memory module called MEMTOK. MEMTOK is SAM's lexicon and general repository of world knowledge (as required by the scripts in the system). MEMTOK converts the conceptual descriptions in the Analyzer output into inferences to real-world objects, places, and people. It also selects lexical items that provide a base surface name for the generative subroutines to express. So when a given script is activated, MEMTOK is instructed by the Applier to use lexical items which are suitable to the situation. For example "chair" occurs in both the restaurant and bus contexts. MEMTOK selects the English realization "chair" in the restaurant situation because a chair is presumably moveable, and "seat" in the bus situation because a seat is presumably non-movable.

The Applier has three basic functions: locating a new input in the data base of scripts, setting up expectations about likely inputs to follow, and instantiating the segments of the script up to the point referenced by the input. Each script in SAM defines a context consisting of:

- patterns which predict what script-level inputs will be seen at a given point in the story;  
- a binding list which links the tokens for objects produced by MEMTOK with script variables;  
- a record of the script scenes currently active;  
- a list of scriptal interferences—i.e., anomalies—that are currently outstanding; and  
- a "strength indicator" which SAM uses to flag how strongly it "believes" in its inferences.

The Applier runs a pattern matcher to decide which of its scripts is being referenced by an input. The two primary features used in this process are the conceptual class an item belongs to—e.g., human, animal, physical object, organization, etc.—and any indicator of the function the item might have—occupation (for persons), title (for organizations), or type (for physical objects). Once an input has been located in a script context, the Applier links it up with what has been previously analyzed in that context, then checks for contradictions. When the linking process has been completed, the Applier updates its predictions about the context based on the new input and prior results. The

---

6 Note that the predictions made by the Applier are at the script level, not the lexical level at which the English Analyzer operates. Script-level predictions are used to aid analysis of the input text; as such, they are a good example of top-down processing. In "Conceptual Dependency: A Theory of Natural Language Understanding," Roger Schank shows how the top-down processing used in CD can resolve a number of comprehension problems, including semantic ambiguity (as demonstrated by the sentence, "The old man's glasses were filled with sherry") and syntactic ambiguity (as demonstrated by the sentence, "Visiting relatives can be a nuisance"). At least seven categories of predictions are discussed by Schank: sentential, conceptual, contextual, conversational, worldview, memory structure, and functional (1972: §7.1 & §7.2).
updated context is then stored and the next round of text processing is started by a call to the English Analyzer.

It should be obvious from the above that processing control moves among the English Analyzer, script Applier, and MEMTOK in a coroutine rather than subroutine manner. In other words, one component may run for a while, send the output elsewhere for further analysis, then regain control and continue processing (incorporating the outcome of the intermediate stage). This kind of operation has definite advantages, not the least of which is the ability to quickly and accurately establish a context that shapes the simulation process.

**The Summarizer** This component of SAM is much less complicated than the previous two. “A good summary must provide two things. It must give the important actions in the prose being summarized; and it must also supply sufficient setting information so any reader will have the necessary context to understand the summary” (Schank and Abelson 1977: 185).

Each scene in a script contains a main conceptualization (MAINCON) from which the eventual summary will be constructed. The Summarizer picks out conceptualizations that are “interesting,” an evaluation made by the following two rules:

1. Interest is judged with respect to the characters in the story. When the Summarizer starts processing a script, it identifies the main characters and their relative importance. For instance, a helpful (if somewhat morbid) heuristic about calamitous events is that the people who were killed are more interesting than the people who were injured, who in turn are more interesting than those who escaped unscathed.

2. Events are interesting if they refer to certain state changes or state values above a given threshold. Thus HEALTH(-1) might not be interesting while HEALTH(-3) could be. These threshold parameters can be manipulated to produce a longer or shorter summary.

The output of the Summarizer is a combination of Conceptual Dependency structures and punctuation commands, which are relayed to the English Generator for natural language expression (see below).

**The Question-Answer Mechanism** The Question-Answer mechanism designed for SAM was oriented specifically toward script-type databases. First, the Analyzer translates English questions into CD representations, which are then fed into the script Applier to generate the proper inferences (also in Conceptual Dependency). Next, inferences in CD are passed on to the Question-Answerer, which also produces its output in CD and must use the Generator for translating answers into English. These operations are summarized in Figure 5.1 (adapted from Lehnert 1977).

The Question-Answerer was motivated by a desire to simulate natural human responses rather than produce adequate but computer-like replies. This required techniques in which the question-answer process consulted information other than that which is explicitly encoded in the story representation. One place where SAM has to look beyond the explicitly stated story is in

---

7 In the context of episodic scripts, MAINCONS are essentially the central events in the episode. In one of their experiments (later discussed in more detail), Bower et al (1979) collected norms which they used to identify the MAINCONS of a number of different scripts.

8 For a more detailed discussion of the processing theory involved see Lehnert (1977).
answering "why not" questions. Suppose John orders a hamburger, is told by the waitress that they do not have any hamburgers, and so orders a hot dog instead. Now we ask, "Why didn't John eat a hamburger?" An appropriate response to this question will be something like, "Because they didn't have any," or "Because the waitress said they didn't have any." But the problem with "why not" questions is that they ask for the causal factors behind events that did not occur, whereas SAM encodes only information about events that did occur. Clearly, SAM must somehow go beyond what is strictly in the input to answer questions of this kind.

Schank and Abelson indicate that "why not" questions can be treated effectively if they are processed as a failure of some anticipated event to come about.

Why-not questions which can be answered are actually questions about failed expectations. "Why didn't John eat a hamburger?" can be answered only because at some time during the understanding process we had an expectation that John would eat a hamburger. This expectation was created when John ordered a hamburger, and it was violated and revised when the waitress told him that they didn't have any. (1977: 188)

To answer questions about failed expectations, points in the storyline where expectations were revised must be identified. This is a fairly straightforward operation for SAM: during input processing, SAM marks such places as points of scriptal interference. A scriptal interference is any event that obstructs the flow of a generic script or script scene. But the word "interference" does not connote an insuperable impediment. In fact, sometimes the solution of an interference is so practiced that it becomes a part of the script itself—e.g., "make another selection from the menu if your first choice is unavailable;" other times the resolution can be trickier, requiring a bit more imagination—e.g.,

---

**Figure 5.1**: Overview of SAM's Question-Answerer. The Analyzer translates English questions into Conceptual Dependency representations. When the Analyzer has translated a sentence, the script Applier is called to generate inferences. The script Applier structures its inferences into a causal chain which can be used by the Question-Answerer. The output produced by the Question-Answer Mechanism is in CD and must be given to the Generator for translation into English.
John’s discovery at Leone’s that he could only pay the check by washing dishes (because his pocket was picked on the subway), a kind of BARGAIN FAVOUR with the management. SAM is able to tag such occurrences and use them to indicate why certain expectations failed to materialize. The program also uses scriptal interferences to answer questions of the form, “Did anything strange or unusual happen . . . ?” The idea is that unusual or unexpected circumstances (of which the obstructions to scriptal flow are a variety) “stand out” in SAM’s memory.\footnote{It should come as no surprise—given the way SAM was constructed and the motivation behind the (script) theory upon which the program was based—that a similar phenomenon seems to operate in human memory as well: generally, the more unexpected or unusual an event is, the more likely it is to be recalled. This effect has been well documented in psychology and given a number of special names associated with the different contexts in which it appears: e.g., the von Restorff effect, (in word recall tasks), flashbulb memory (in connection with the affects of emotions on LTM storage), and the salience or vividness bias (related to the availability heuristic used in judgement and reasoning). See Ashcraft (1994) for further details and discussion of the relevant literature.}

SAM now constructs a “ghost path” based on the failed expectation (Lehnert 1977: 64). That is, having been told that John ordered a hamburger, SAM creates a phantom event sequence which predicts the waitress taking the order, the cook preparing the hamburger, John eating it, and so on. Answers to “why not” questions are generated by tracing the ghost path to the point of interference. Thus “Why didn’t John eat a hamburger?” is answered by “Because the waitress told him they didn’t have any.”\footnote{Similarly in Leone’s, “Why couldn’t John pay the check?” is answered by “Because John did not have any money;” and “Why didn’t John have any money?” is traced back to “Because a thief picked John’s pocket.”}

The English Generator Goldman’s English Generator program—called BABEL—has been incorporated in SAM. BABEL handles input of Conceptual Dependency and produces English sentences as output. We will not go into the operational details of this component here, but those interested should consult Goldman’s paper, “Conceptual Generation.”\footnote{There is a sixth component of SAM that we will also ignore—namely, a Chinese Generator (or translator) which is a modified version of BABEL.}

ii. FRUMP

Like SAM, FRUMP (Fast Reading, Understanding, and Memory Program) is a script-based system. But unlike SAM, FRUMP was designed to skim a given text quickly and extract from it only the most important information for the purpose of producing a brief summary. Thus FRUMP does not engage in the kind of complicated analyses that depend on making predictions and drawing inferences; and for this reason, the system has no need for the sort of detailed scripts to which SAM had access, but uses instead sketchy scripts.

The crucial difference is that sketchy scripts have far fewer Conceptual Dependency representations (only those corresponding to the most important events in SAM’s scripts) and more often than not, the causal connections between conceptualizations are not included. The result is that FRUMP understands most of what is important to understand in news articles and works very much faster than SAM. (Schank and Abelson 1977: 204)

For each type of newspaper story, FRUMP has a list of expected facts—called requests—which it seeks to find. Requests are in CD format and contain unfilled slots. In processing an article, FRUMP must select a sketchy script and then try to find in the article occurrences of the facts
called for by the requests. When an instance of one of the requests is found and its slots filled, that request is said to be satisfied.

FRUMP is composed of three distinct components: a Parser, a script Applier, and an English Generator. The Parser employed is phrase (rather than word) oriented. It parses phrases from the input text into Conceptual Dependency representations. The script Applier then matches these CD representations against the program’s repertoire of scripts. When a match is found, the slots in the requests are filled. This continues until all the requests are satisfied, at which time the outcome is fed into the Generator to be expressed as a brief English summary. Here is an example.

Input: Officials here said today that the death toll from the collision of an excursion train with a freight train on Sunday has risen to 23; 45 people were injured. The dead include two Americans. Responsibility for the accident was laid to the passenger train engineer who apparently failed to heed a stop signal.

The district attorney’s office said two other Americans had been injured in the collision near the Barranca Del Cobre in the Sierra Madre.

All the dead were Mexicans except the two Americans and two Britons. Most of those aboard were Mexican travel agents.

The Americans were identified as Mark Mortellaro of New York and Martin Ward whose hometown is still unavailable. The district attorney’s office said the two injured Americans were Paul Joseph Callsen and Mary Callsen, both of New York.

Output: A TRAIN HIT A TRAIN IN MEXICO. 23 PEOPLE DIED. 45 PEOPLE WERE INJURED. THE ENGINEER WAS BLAMED FOR THE CRASH.

If the output seems curt and somewhat lacking in informational content, it should be remembered that both the “sketchiness” of the script and the number of requests could be manipulated to produce summaries of various lengths and detail. It should also be remembered that FRUMP is not like SAM. The former system was designed to skim stories in the same casual manner a person reading a newspaper might, taking in only the most pertinent information, not analyze text in any significant depth.

iii. TALESPIN

TALESPIN is a program designed to create a simulated world, assign goals to the characters in that world, and then indicate what happens when these goals interact with various events. TALESPIN develops stories by violating a goal state for one of its characters and then creating a plan to achieve the frustrated goal. As such, the entire apparatus of scripts, goals, and planboxes is used by the program. An interesting departure from the previous systems is that the user is expected to interact with TALESPIN, making choices of characters, relationships, and story locations.

The following is an example run of TALESPIN. The output column shows the program’s reports and prompts for action from the user; the input column shows the user’s choices; the comments column contains some explanatory remarks.

12 The program is explained in reasonable detail by Schank and Abelson (1977: 210-17).
<table>
<thead>
<tr>
<th>Output</th>
<th>Input</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHOOSE ANY OF THE FOLLOWING CHARACTERS FOR THE STORY:</td>
<td>Bear, Bee, Boy, Canary.</td>
<td>The program first prompts the user to choose the characters in the story. When the characters are selected, TALESPIN starts to create a world for them.</td>
</tr>
<tr>
<td>BEAR, BEE, BOY, GIRL, FOX, CROW, ANT, HEN, LION, DOG, WOLF, MOUSE, CAT, GOAT, CANARY.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JOHN BEAR IS AT A CAVE. A BEEHIVE IS AT AN APPLE TREE. IRVING BEE IS AT THE BEEHIVE. SOME HONEY IS AT THE BEEHIVE. IRVING BEE HAS THE HONEY. SAM ADAMS IS AT A HOUSE. WILMA CANARY IS AT A NEST.</td>
<td></td>
<td>These facts are now added to memory, along with related inferences—e.g., that the characters know where they are, what they possess, what they need, etc.</td>
</tr>
<tr>
<td>CHOOSE ANY OF THE FOLLOWING ITEMS: BERRIES, FLOWERS, RIVER, WORM.</td>
<td>Berries, Worm.</td>
<td>TALESPIN continues to develop the storyline by asking the user to choose some items. The list from which the user must choose is not arbitrary. The program selects items that are somehow related to the characters involved. Hence, &quot;Worm&quot; would probably not have appeared had &quot;Canary&quot; not been chosen earlier by the user.</td>
</tr>
<tr>
<td>SOME BLUEBERRIES ARE AT A BUSH. A WORM IS AT A PATCH OF GROUND.</td>
<td></td>
<td>The program situates the chosen items, and now asks some questions in order to create a network of relationships between the characters.</td>
</tr>
<tr>
<td>WHO KNOWS ABOUT THE BLUEBERRIES?</td>
<td>4</td>
<td>This is the satisfaction of D-KNOW for John Bear.</td>
</tr>
<tr>
<td>1. WILMA CANARY</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. SAM ADAMS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. IRVING BEE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. JOHN BEAR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JOHN BEAR THINKS THAT THE BLUEBERRIES ARE AT THE BUSH. HOW HUNGRY IS JOHN BEAR?</td>
<td>4</td>
<td>TALESPIN's BEAR85s contains the information that bears are stereotypically fond of honey. Since there is some honey at the beehive, if it turns out that John Bear is hungry, the program can setup a goal for that character involving a D-CONT. But the user indicates that John Bear is not hungry at all, so TALESPIN has to try something else.</td>
</tr>
<tr>
<td>1. VERY</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. SOMEWHAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. NOT VERY</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. NOT AT ALL</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>WHO KNOWS ABOUT THE WORM?</strong></td>
<td>2</td>
<td>The worm-bird is another path for TALESPIN to try, so it asks who knows about the worm. If the input had been 1, the next question would have been: &quot;HOW HUNGRY IS WILMA CANARY?&quot; But the user chooses 2.</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>1. WILMA CANARY</td>
<td>2. SAM ADAMS</td>
<td>3. IRVING BEE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>SAM ADAMS THINKS THAT THE WORM IS AT THE PATCH OF GROUND.</strong></th>
<th></th>
<th>Since Sam Adams does not eat worms (and therefore has no reason to want to acquire any), this path seems to be a dead-end (at least for now).</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th><strong>THIS STORY IS ABOUT ...</strong></th>
<th>2</th>
<th>Sam Adams is chosen as the main character of the story.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. WILMA CANARY</td>
<td>2. SAM ADAMS</td>
<td>3. IRVING BEE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>HIS PROBLEM IS THAT HE IS ...</strong></th>
<th>1</th>
<th>The main character gets &quot;a problem.&quot; It is not necessary that this problem be solved, nor need it be the main focus of the story (though it will be in this example). The problem simply acts as a starting point from which the plot can develop.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. HUNGRY</td>
<td>2. THIRSTY</td>
<td>3. HORNY</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>SAM ADAMS IS HUNGRY. SAM ADAMS WANTS TO GET SOME BERRIES.</strong></th>
<th></th>
<th>Sam is hungry, so TALESPIN sets up a goal for him to satisfy his hunger. It checks among the item list to see if anything there can do the job, and comes up with blueberries.</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th><strong>DOES SAM ADAMS LIKE WILMA CANARY?</strong></th>
<th>2</th>
<th>Sam Adams wants the berries, but one cannot get a thing without knowing where it is. One way to satisfy D-KNOW for Sam is the ASK planbox. Since ASK works better when the askee is a friend, TALESPIN wants to find out who among the characters in the story is friendly to Sam. Wilma canary seems to fit the bill, but other aspects of their relationship still need clarification.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. A LOT</td>
<td>2. A LITTLE</td>
<td>3. NOT MUCH</td>
</tr>
<tr>
<td>Question</td>
<td>Score</td>
<td>Answer</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------</td>
<td>-------</td>
<td>--------</td>
</tr>
<tr>
<td>DOES SAM ADAMS FEEL DECEPTIVE TOWARDS WILMA CANARY?</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>1. A LOT 2. A LITTLE 3. NOT MUCH 4. NOT AT ALL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOES SAM ADAMS FEEL COMPETITIVE TOWARDS WILMA CANARY?</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>1. A LOT 2. A LITTLE 3. NOT MUCH 4. NOT AT ALL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAM ADAMS DECIDES THAT WILMA CANARY MIGHT WANT SAM ADAMS TO GIVE WILMA CANARY A WORM.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAM ADAMS WANTS TO ASK WILMA CANARY WHETHER WILMA CANARY WILL TELL SAM ADAMS WHERE SOME BERRIES ARE IF SAM ADAMS GIVES WILMA CANARY A WORM.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAM ADAMS WALKS FROM THE HOUSE TO THE GROUND BY THE REDWOOD TREE BY GOING THROUGH A MEADOW. SAM ADAMS IS AT THE GROUND BY THE REDWOOD TREE.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAM ADAMS ASKS WILMA CANARY WHETHER WILMA CANARY WILL TELL SAM ADAMS WHERE SOME BERRIES ARE IF SAM ADAMS GIVES WILMA CANARY A WORM.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOES WILMA CANARY FEEL DECEPTIVE TOWARDS SAM ADAMS?</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1. A LOT 2. A LITTLE 3. NOT MUCH 4. NOT AT ALL</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Since the user indicates that Sam feels competitive (but not deceptive) towards Wilma Canary, TALESPIN will not have Sam Adams use the ASK planbox: there are constraints related to when and how a given planbox may be used, and one condition on the use of ASK is that the players involved should not be in competition. TALESPIN must consider another possibility in the persuade package.**

**TALESPIN has decided on BARGAIN OBJECT. The program's CANARYs contains the information that birds eat worms. So it is reasonable to offer to bring Wilma Canary a worm in exchange for information about the whereabouts of the blueberries.**

**Before Sam can talk to Wilma, he has to undertake a D-PROX. Then he makes his offer.**

**In order to know how Wilma should answer, TALESPIN solicits information about Wilma's perception of her relationship with Sam. The user makes Wilma a very sly bird!**
| WILMA CANARY TELLS SAM ADAMS THAT WILMA CANARY WILL TELL SAM ADAMS WHERE SOME BERRIES ARE. | Wilma Canary agrees to Sam’s proposal, but we know that Wilma is not to be trusted. Sam has no such knowledge, however, so he decides to get the worm, which requires him to undertake a D-PROX. Sam returns to Wilma Canary with the worm (another D-PROX) and hands the gift to her. |
| SAM ADAMS WANTS TO GET NEAR THE WORM. SAM ADAMS WALKS FROM THE GROUND BY THE REDWOOD TREE TO THE PATCH OF GROUND BY GOING THROUGH THE MEADOW. | |
| SAM ADAMS TAKES THE WORM. | |
| SAM ADAMS WANTS TO GET NEAR WILMA CANARY. SAM ADAMS WALKS FROM THE PATCH OF GROUND TO THE REDWOOD TREE BY GOING THROUGH THE MEADOW. SAM ADAMS GIVES WILMA CANARY THE WORM. | |
| HOW HUNGRY IS WILMA CANARY? | 1 |
| 1. VERY | As soon as Wilma gets the worm, the user is asked about the bird’s state of hunger. |
| 2. SOMEWHAT | |
| 3. NOT VERY | |
| 4. NOT AT ALL | |
| WILMA CANARY EATS THE WORM. THE WORM IS GONE. WILMA CANARY IS NOT VERY HUNGRY. | All traces of the worm are now erased from TALESPIN’s memory. |
| WILMA CANARY TELLS SAM ADAMS THAT SAM ADAMS IS NOT AT ALL SMART. | Wilma tricks Sam and adds insult to injury. In consequence, Sam’s feelings toward Wilma have understandably changed. |
| SAM ADAMS LIKES WILMA CANARY NOT AT ALL. | |
| DOES SAM ADAM DOMINATE WILMA CANARY? | 2 |
| 1. A LOT | A somewhat ominous question that signals an unfortunate shift in Sam’s tactics. The user gives an affirmative answer. |
| 2. A LITTLE | |
| 3. NOT MUCH | |
| 4. NOT AT ALL | |
| SAM ADAMS STRIKES WILMA CANARY. WILMA CANARY IS NOT AT ALL HEALTHY. WILMA CANARY WON’T TELL SAM ADAMS WHERE SOME BERRIES ARE. | OVERPOWER, it seems, is Sam’s new strategy. But although Wilma pays for her deception, she remains steadfast! |
| SAM ADAMS DIDN’T FIND OUT WHERE SOME BERRIES ARE. | Sam ran out of things to try. And since his satisfying his hunger was the main goal around which the story revolved, the resolution of this goal, whether positively or (in this case) negatively, marks the conclusion of the tale. |
| SAM ADAMS DOESN’T GET SOME BERRIES. | |
| THE END | |
Chapter V  Scripts II: Evidence from AI and Psychology

**ELIZA** It should be noted that the programs described above have not been rigged with any tricks that made the computer appear more "intelligent" than it actually is, contrary to what has been done in a number of other cases. One example is ELIZA—so called because, "like the Eliza of Pygmalion fame, it can be made to appear ... more civilized, the relation of appearance to reality, however, remaining in the domain of the playwright" (Weizenbaum 1966: 36). The program was designed by Joseph Weizenbaum to carry out a conversation with a human subject in a natural manner. But unlike SAM, ELIZA was not scriptally structured.

ELIZA functions in the following way: the input text is read and compared against a store of keywords. If matches are found, the sentences containing the keywords are analyzed on the basis of decomposition rules, and responses are generated by reassembly rules associated with the rules of decomposition. The main idea is one of text manipulation, at the heart of which are transformation rules which serve to take apart and then reassemble the data string according to fixed criteria. All of this is carried out without regard to context.

Consider the sentence "I am very unhappy these days". Suppose a foreigner with only a limited knowledge of English but with a very good ear heard that sentence spoken but understood only the first two words "I am". Wishing to appear interested, perhaps even sympathetic, he may reply "How long have you been unhappy these days?" What he must have done is to apply a kind of template to the original sentence, one part of which matched the two words "I am" and the remainder isolated the words "very unhappy these days". He must also have a reassembly kit specifically associated with that template, one that specifies that any sentence of the form "I am BLAH" can be transformed to "How long have you been BLAH", independently of the meaning of BLAH. A somewhat more complicated example is given by the sentence "It seems that you hate me". Here the foreigner understands only the words "you" and "me"; i.e., he applies a template that decomposes the sentence into the four parts:

\[
\begin{align*}
(1) & \text{ It seems that} \\
(2) & \text{ you} \\
(3) & \text{ hate} \\
(4) & \text{ me}
\end{align*}
\]

of which only the second and fourth parts are understood. The reassembly rule might then be "What makes you think I hate you"; i.e., it might throw away the first component, translate the two known words ("you" to "I" and "me" to "you") and tack on a stock phrase (What makes you think) to the front of the reconstruction. (Weizenbaum 1966:37)

When there are no keywords on which ELIZA can latch, the system spews out what Weizenbaum calls content-free phrases: "Please go on," "That's very interesting," "I see."

In order to keep the user in the dark about the true nature of his or her conversation partner, Weizenbaum adopted a "psychiatric interview" format for the exchange. He rationalizes his choice by observing that "if ... one were to tell the psychiatrist 'I went for a long boat ride' and he responded 'Tell me about boats', one would not assume that he [the psychiatrist] knew nothing about boats, but that he had some purpose in so directing the subsequent conversation" (1966: 42). This kind of trickery has the advantage of eliminating the need for ELIZA to store any explicit information about the real world; it is hoped, rather, that "the human speaker will contribute much to clothe ELIZA's responses in vestments of plausibility" (1966: 42).

The difference between ELIZA and SAM is not hard to see. ELIZA never really goes beyond the input text; it simply decomposes and reassembles sentences as dictated by static guidelines. The program neither represents the world nor uses explicit information to assist its performance. SAM, on the hand, was designed precisely to form schematic representations (in the form of scripts) and then use these schemas to generate responses. This makes SAM exceptionally context-sensitive; it has the ability to go beyond the input text by tracing causal connections, drawing inferences, making predictions, and answering questions about events that were not mentioned in the storyline. ELIZA can't do any of this, and no amount of improvement (short of a comprehensive overhaul) is likely to
produce better results. The reason has to do with the limitations of the basic approach behind ELIZA's design: the program was not created to simulate human interactions; it was created to mimic human responses.13

iv. Modelling Cognition

Let me close by making some observations about the current status of script theory in both AI and psychology, followed by a few words about the relationship between the script-based computer modelling described above and human cognition.

Script theory is no longer actively pursued in AI research, and has not been so for at least twenty years. The main reason for this has to do with issues relating to practical application, not with the legitimacy of underlying notions or presuppositions: humans have a huge repertoire of scripts that provide an astronomical amount of background information; this information helps us deal with and understand common situations and ordinary encounters. Trying to implement these scripts in a computer program in a manner that reasonably reflects the extent of human knowledge proved realistically unfeasible; so script theory was abandoned.

Even so, practical applicability—AI's overriding condition for sustained research interest—is not the criterion to be used in assessing whether or not a conceptual entity is helpful, or has explanatory potential, or otherwise deserving of consideration. That is because the research aims of AI tend to be much more sensitive to pragmatics than the research aims of psychologists and philosophers of mind. It is perhaps not very surprising, therefore, that scripts are still alive and well in psychology. Although the word "script" is not always used,14 the central idea—the idea of a schema—has been around for more than seventy-five years and shows no signs of waning. In fact, schemata, going by various names, permeate current psychological theorizing at a very deep level. One can hardly pick up a cognitive (or social) psychology textbook without finding numerous references (both implicit and explicit) to some sort of stable, highly organized human knowledge structures that function in exactly the same manner in which scripts have been envisioned to function.

My hypothesis is that some Intentional states are based on scripts; I say which in chapter IV and I give some indication of how these scripts are structured. The evidentiary value of computer simulation is derived from the manner in which a scriptal organization of information allows script-based programs to carry out certain tasks. Both the script model and the human agent solve simple problems of a specific type, so that script theory provides some insight into the manner in which information can be structured in order to solve these problems. Newell (1973: 27) calls this the sufficiency test, which is highly characteristic of the information-processing conception of AI.

On the other hand, none of this says anything about the content of thought. Obviously, SAM gets the content of its scripts from human programmers, but from where do the programmers get the content of their scripts? Giving an account of the semantics of human Intentionality is the business of the next chapter. And until this is done, it will not be clear just how scripts can form the

13 I once played a Turing-test game with a computer program that had a design similar to ELIZA's. I distinctly remember a lot of please-go-on's and I-see's, many of which seemed out of place. Needless to say, the program's performance was not very inspiring.

14 See point (viii) in §1.4.
foundation for a view that is sensitive to the first-person perspective. Nevertheless, what we are interested in here is how well an AI program functions once scriptally-organized information has been "installed" in it. We are concerned, that is, with the information-processing advantages or capacities that a scriptal organization of data brings, not the source of the content or meaning of this data. It is not a strong AI thesis in Searle’s (1980) sense that I intend SAM and the other script-based programs to support. Nor is it necessary for anyone to claim that computers must have real language understanding in order to illuminate the way in which contentful information may be structured in human memory in order to enable certain kinds of cognitive operations.

Computer modelling can (and does) tell us how well certain knowledge structures work, their advantages and disadvantages, and in this way gives researchers the opportunity to develop and perfect their hypotheses. SAM and the other script-based programs provide evidence for the operation of scripts in human cognition because of all the things scripts enable the programs to do: summarize passages, answer questions about stories, construct imaginary worlds with characters that have goals and plans, draw inferences, and plot (in)direct causal relations in event-sequences. Is this kind of evidence absolute or definitive? Certainly not. But neither is it insignificant.

5.2 Support from Psychology

The first part of this section outlines some of the evidence for the taxonomic division of long-term memory into the episodic, procedural, and semantic components. One of the claims I made back in §1.4 is that scripts are not (necessarily) linguistic models. The findings reported below are meant to support this assertion. The fact that many scripts are non-linguistic is verified by studies showing that the LTM components housing the different kinds of script-based knowledge are dissociable from each other and from the semantic system. Researchers have begun to track down the neural structures involved in LTM using a combination of methods, including investigating patients with neurological damage and brain imaging techniques.

i. The Semantic-Episodic-Procedural Dissociation in LTM

Psychologists define declarative memory as encompassing the acquisition, retention, and recall of past events (episodic memory) and of generic and linguistic facts (semantic memory). In contrast, non-declarative memory encompasses the acquisition, retention, and retrieval of the kind of information that is expressed through experience-induced changes of performance; that is, the non-declarative system encompasses what we have called procedural knowledge. There is good evidence that the declarative and non-declarative long-term stores are dissociable, that procedural memory is distinct from both episodic and semantic memory.

One of the ways researchers discovered the distinction between declarative and non-declarative memory was by observing patients who suffered from an anterograde memory deficit—a kind of "forward" amnesia which prevents those who suffer from it from learning new declarative data. Anterograde amnesia is symptomatic of certain kinds of brain lesions caused by accident or disease. Although individuals diagnosed with this condition experience difficulty in sorting and retrieving new semantic and episodic information, they show minimal impairment on procedural tasks. This was demonstrated dramatically in a famous (1968) case study by Brenda Milner and her colleagues.
A man identified only as HM underwent surgery to control life-threatening epileptic seizures. The surgeon removed most of HM's medial temporal lobes, including the hippocampus and amygdala (see Figure 5.2). Following the operation, HM had one of the deepest cases of amnesia ever recorded. Although he was free of seizures, he lost the capacity to encode new declarative information in memory: every time he met Dr. Milner, who studied him for over 20 years, he had to be reintroduced to her, and would then smile politely and tell her that it was a pleasure to make her acquaintance. But despite HM's impairment, he had no difficulty in learning new procedural skills, such as writing words upside-down.  

Figure 5.2: The medial temporal region (inside the middle of the temporal lobes), particularly the hippocampus, plays a key role in the processing of semantic (or declarative) information. The frontal lobes play a more important role in working memory, procedural memory, and some aspects of episodic memory (such as dating events for the time at which they occurred). Posterior regions of the brain (occipital cortex, parietal cortex, and temporal cortex) are involved in processing sensory data and the production of mental representations.

Over the last two decades, lesion research has shown that the hippocampus and adjacent anatomically related regions of the cortex are central for the storage and retention of declarative information (Gluck and Myers 1997). The fact that amnesiacs like HM often demonstrate normal skill-learning capacity suggests that the hippocampus, while important for the consolidation of declarative information, is not central to procedural memory. In fact, in some cases the dissociation between procedural and declarative knowledge has been remarkably specific: some patients with cortical lesions have shown selective deficits for retrieving the names of (a) people and other proper nouns (Semenza and Zettin 1989), (b) fruits and vegetables (Hart et al. 1985), (c) living things such as animals (Damasio et al. 1996), and (d) manufactured objects like tools (Damasio et al. 1996).

Interestingly, the distinction in LTM that was long supported by work with victims of neurological damage is that between episodic declarative memories and procedural knowledge. And indeed, we observe just this dissociation in HM who was unable, not only to encode semantic information like the names of his caregivers, but also memories of occurrences: he was never able to store the event of meeting Dr. Milner (or any other new acquaintance, for that matter), and although he learned how to write words upside-down, each new time he was asked to do so he had no recollection that he had ever performed the task before. Yet the reported improvements in his speed

Better known as that “mirror tracing” task, where subjects trace letters reflected from a mirror, seeing only their hands and the writing instrument.
and accuracy indicate that he had somehow retained the skill, but not the corresponding learning episodes.\footnote{Cohen and Squire have conducted similar experiments with other amnesiacs and have reached similar conclusions. Amnesic patients acquire a mirror-reading skill at a rate equivalent to matched control subjects and retained it for at least 3 months, they report. “The results indicate that the class of preserved learning skills in amnesia is broader than previously reported. Amnesia seems to spare information that is based on rules or procedures, as contrasted with information that is data-based or declarative—‘knowing how’ rather than ‘knowing that.’ The results support the hypothesis that such a distinction is honoured by the nervous system” (1980: 207). And J. D. E. Gabrieli observes, “amnesic patients gain skill in reading such text at a normal rate, despite poor declarative memory for the particular words read or the episodes in which they gained their skill” (1998: 99).}

An interesting finding has recently surfaced in mirror-reading procedures. Using functional magnetic resonance imaging (fMRI), Poldrack et al. (1998) found that after practicing reading mirror-reversed words for a period of time, normal subjects showed decreased activity in their visual neural pathways but increased activity in their verbal neural pathways. This suggests that the subjects were moving from the visual skill of mentally rotating the word to the linguistic skill of understanding the word’s meaning—showing, once more, that the two tasks take place in different regions of the brain.

These studies, along with countless others, point to a fairly well defined separation of the declarative (episodic, semantic) and non-declarative (procedural) memory systems. But what about the episodic and semantic stores? Since Tulving first enunciated the episodic-semantic distinction, evidence for the cleavage has been accumulating. In a 1989 paper entitled “Remembering and Knowing the Past,” he discusses two kinds of retrieval processes in human memory. Her refers to one kind as the retrieval of episodic information, or simply episodic retrieval, and to the other as the retrieval of semantic information, or simply semantic retrieval. Tulving also employs the terms “remembering” or “recollecting” for episodic retrieval, and “knowing” and “recalling” for semantic retrieval. Using this terminology, he puts forth the claim that “remembering one’s past is a different, perhaps more advanced, achievement of the brain than simply knowing about it” (1989: 367). He then outlines two lines of evidence showing that episodic and semantic memories really are different components of LTM.

The first kind of evidence is based on another case history, this time of a patient known as KC, who, because of a serious motorcycle accident resulting in brain injury, displayed a rather unique form of amnesia. The following is an excerpt from Tulving’s description of the extent of KC’s memory problems:

KC’s case is remarkable in that he cannot remember, in the sense of bringing back to conscious awareness, a single thing that he has ever done or experienced in the past ... Those aspects of KC’s intellectual functioning that do not depend on remembering personal experiences are reasonably normal. His measured IQ is within the normal range ... he recognizes familiar objects ... his understanding and use of language are unimpaired ... and his thought processes are intact ... But KC does not remember any personally experienced events from either before or after his accident ... Here are a few examples. KC knows that his family owns a summer cottage, knows where it is located, and can point out the location on a map of Ontario, and he knows that he has spent summers and weekends there. But he does not remember a single occasion when he was at the cottage or a single event that happened there. He has retained his knowledge of how to play chess, but he cannot remember having played chess ever before, with anyone. He can only guess that he played with his father, because he knows that his father plays chess. He knows that he owned a car and can recall its make and year. But he cannot remember a single trip he took in the car. He knows, and can describe in great detail, the exact sequence of steps to be taken—the “script”—when changing a flat tire on a car. But he cannot tell whether he himself ever had to change, or witnessed the changing of, a flat tire, since he does not recollect any such occasion from his life. For the three years immediately preceding his accident, KC worked for an engineering company. He knows that he did so, and
he can recall the name of the company and the nature of its business. But he does not remember working there: he cannot provide a description of his workplace, and he does not recognize a color photograph of the office on the factory floor that he occupied for three years. Finally, he knows the meanings of technical terms such as "spiral mandrel" and "extruder screw," terms that he learned in the course of his work. But he does not remember a single event, or even any repeated events, that happened during that time. (1989:362-3)

These kinds of contrasts between what KC does not remember of his past and what he knows from it support the idea that episodic memory and semantic memory are subserved by different neural mechanisms. The kind of brain damage KC suffered in his motorcycle accident seems to have resulted in serious impairment in the functioning of his episodic memory system, but little impairment in his semantic (and procedural) system. This pattern of deficit supports the idea that episodic and semantic memory are separate mechanisms. But as Tulving himself acknowledged, there are limitations on what can be learned about normal cognition from data obtained from brain-damaged patients. KC is unique. Very few other reports describe a person with amnesia so severe as to prevent him from remembering anything about his personal past. Because psychologists are dubious about the generality of such single-case results, Tulving presents further evidence for his position.

This second line of evidence involves measuring the blood flow in the brains of normal subjects under controlled conditions. The logic behind this procedure is that a mental task produces an increase in neural activity (relative to a known baseline), which should be manifested as a rise in blood flow to those region(s) in the brain housing the increased activity. Thus by injecting a dose of radioactive material (irradiated gold, in Tulving's case) into the bloodstream of a subject, a positron emission tomography (PET) scanner can measure the areas of the subject's brain that have higher concentrations of radioactivity when (s)he is asked to perform a given task. In this way, different brain structures can be plotted and associated with various types of cognitive functions.

In Tulving's study, subjects were asked to retrieve various semantic and episodic memories. In one case, the semantic component was a request to retrieve the contents of a previously-read history of astronomy book (discussing the theories and achievements of Copernicus, Brahe, Kepler, and others), while the episodic component involved the same subject recalling personal memories from a summer nearly 50 years earlier. The results of these experiments were clear: different brain activities occurred when the two kinds of memories were retrieved. The same finding was obtained when subjects were asked to retrieve recent semantic and episodic memories. In particular, episodic retrieval was accompanied by greater activation in the anterior structures of the brain, while semantic retrieval appears to activate more posterior structures.

Tulving's studies have recently been duplicated and corroborated by Lars Nyberg, who observes that PET technology has provided "evidence at the neural level that different episodic memory tasks share component processes, and that these processes are not tapped by non-episodic memory tasks. As such, PET studies of episodic memory have provided important converging evidence for the separation of various long-term memory systems" (1998: 112).

---

17 And have you noticed the dissociation between KC's semantic and procedural memory in Tulving's description?
18 But see Caramazza (1986) for a spirited defence of single-patient studies.
ii. Schemata and Scripts

In *Scripts, Plans, Goals, and Understanding*, Schank and Abelson took the position that long-term memory is organized episodically rather than hierarchically. "Not only is information stored by humans in episodic form; it is also acquired that way. For example, a child learns about the order of processing in a restaurant by being dragged through the experience enough times. He learns in the same way about department stores, and only much later does he see any similarity. Thus, we would expect that an important part of the language acquisition process is the acquisition of scripts" (1977: 222). So would I, if only for that subclass of linguistic forms which refer to the objects and situations of ordinary life.

But if the answer to the question "Where do scripts come from?" is that they are acquired through experience, we should be able to identify evidence for the scriptal acquisition process in early childhood, where most of the learning about self and environment takes place. Roger Schank searched for evidence for scripts in children by conducting a study of his own daughter, Hana, at various ages.19 His findings should not be overestimated, however, since the research was purely informal; but I do think it is important and deserves to be mentioned.

At age 3y-4m, Hana was asked about her knowledge of restaurants.20

P. Tell me a story—what happens in a restaurant? What happens—you go inside the restaurant ...
H. You sit down, and you uh, eat food ...
P. How do you get the food?
H. From the waitress.
P. How does the waitress know what to give you?
H. If you ask for a hamburger, then she gives you hamburger.
P. What happens if you ask for hot dog, do you get Hamburger?
H. No you get hot dog.
P. And then what happens after she gives you the food?
H. She gives you dessert.
P. And then what happens?
H. And then you leave.
P. And then you leave? Just like that?
H. No, the waitress gives you some money and you pass some money to her and she gives you some money back to you and then you leave.

The mistakes and omissions in Hana’s story are a reflection of what is important to her or what she finds interesting in the restaurant experience at this age: her RESTAURANT es is concerned mostly with eating, and irrelevancies such as paying the check are beyond her at this point in her cognitive career. Schank’s reports show, however, that Hana’s understanding of standard situations had somewhat improved by age 4y-2m, when the RESTAURANT es was once more elicited from her.

P. Now, I want you to tell me what happens when you go to a restaurant.
H. OK.
P. What happens in a restaurant? Start at the beginning.
H. You come in and you sit down at the table. And then the waitress comes. And she gives you a menu. And, then she takes it back and writes down your order. And, then you eat what she gave you. And, then you get up from the table. And you pay the money and then you walk out of the store.

19 Ages are shown throughout as u years and v months, written uy-vm.

20 "P" denotes Papa and "H" denotes Hana. All dialogues are from Schank and Abelson (1977: 223-30).
It seems that eight months have made quite a bit of difference. Overlooking the “store” misreference in the last sentence, at 4y-2m Hana’s RESTAURANTEs appears to have all the main elements that the script has in adult subjects (but probably with fewer details). Schank and Abelson claim that the process of script acquisition begins with the child assuming that everything she encounters will happen the same way again the next time. The assumption seems sound, since the power to readily differentiate between the common and incidental features of numerically different but type-identical episodes comes much later in cognitive development. So, at least early on, scripts of all kinds are being constantly elaborated and revised with each successive experience.

Hana was also asked to make up stories at various ages. These stories give a glimpse into the process of script generalization and the onset of schematic planning. At 2y-11m, Hana told a story about a bicycle ride:

P. I met you in the plant store and then what happened?
H. I don’t know.
P. Oh come on, I thought you were going to tell me a story.
H. I went on a bike and you drived, and you buckled me in and we went riding the park, and droved and went on one swing and then the other swing and two swings and then we, I went on two slides and then went in the sand box and then I met somebody, and then I met, and then I went right on our bicycle, and took the thing off and and we went on and you droved on and then you taked the one street and then you ride on the other side and then we were home and then we were opened, I opened, you opened, I opened the door and then I went right in and I take my jacket off and said hang up your coat and I hanged my coat and said and said and went right in the kitchen and and was almost time to dinner.

Hana’s story is obviously incoherent and lacking in goal organization. Everything seems to be of equal importance to her, which is exactly the opposite of the purpose and general usage of scripts in adults (Schank and Abelson 1977: 227). But by age 3y-4m, the organizational influence of scripts starts to appear in her storytelling.

P. Tell me a story.
H. I don’t know what happened.
P. Once upon a time ...
H. There was a little girl named Hana Rana Bana and she was riding her hot wheels [this didn't actually happen] and David Shapiro21 was riding a car and they went up and down the driveway into the backyard zoom across, up the concrete.
P. Across what?
H. Papa taked a walk with David Shapiro and Hana ... Rana Bana and while she was riding and she stopped where the stop sign was with David Shapiro and Hana said, David Shapiro, let’s walk that way, they said let’s walk that way, [pointing opposite direction], they said let’s walk that way, so they went that way and soon they caught up and they turned the corner and turned and turned till they went, cause they were going on a real vacation, and they turned and turned and went on and on with Joshua and Hana Rana Bana and David Shapiro and soon they came to Gammy and Poppy’s house and they gave lots of presents even a ruler [she just got a ruler from them] even a drop of drippy drops and then they just were sleeping away at at Gammy’s and Poppy’s house and had dinner and played everything, and then a new day they came to the Concord, [a soon to be taken vacation] they swimmmed, and they played and they had lots of fun. So then after a while, they went home to their own house and they went to sleep, to play, that’s all.

What we have in this case is a joining of two scripts, a PLAY-OUTSIDES and a VACATIONES. (There could also be a third script in there somewhere, “visiting Gammy and Poppy’s house.”) The first contains the idea of going someplace (up and down the block), which happens to call up or prime the vacation

21 An acquaintance of Roger Schank’s.
script. Not everything is equally important anymore, and the events, thought at times a little confusing (e.g., the “drop of drippy drops”), seem a lot more lucid than those told by Hana only five months earlier. We now get a story that is more goal-oriented. And while we can still see just how much her personal experiences contributed to the schematic organization of her story (e.g., the ruler and the Concord), we may notice that they do so in a much more controlled way.

Although Schank’s reports are telling, it is natural to want something a little more formal. In fact, investigations have shown that children as young as 2 and 3 begin to establish reliable and valid representations of the major events in familiar, stereotyped sequential activities—such as eating lunch at nursery school, taking a bath, and participating in a telephone conversation. One such study was carried out by Katherine Nelson and Janice Gruendel, who have scrutinized the role of scripts in children’s conversations. “It is our claim,” they assert,

that the development of topic relevant dialogue structure may profitably be viewed as a function of building up of shared social scripts which specify the structure and content of familiar events in the child’s experience. In fact, the conversational structure may itself be viewed as one kind of very general social script. ... Our recent research on how young children build up scripts ... is based on the belief that they are a basic organizing structure for the young child’s knowledge system. In our view, scripts are useful to the child for predicting the sequence of routines, anticipating who will act and how, recognizing alternative slot fillers, and in general operating as an expert on the passing scene. It is the well-known scripts for baths, feeding and other home activities that enable the child to take an active part in the action as well as in the conversation about the action, even when out of the context of the immediate situation. (1979:78-9)

Nelson and Gruendel found that playing and eating are very general script components for young children (1979: 79-80). If pressed, they will mention what they play and what they eat, but their usual reference to these episodes is the general event term itself, a strategy which leaves open the possibility of later filling in context-specific frames. A second important finding was a high level of agreement on the basic event sequences of the scripts with which children are acquainted. Such commonalities are an obvious support for the usefulness of scripts in social interactions.

Studies with adults have also been undertaken, and some earlier research regarding the structure of schemata—e.g., that published by F. C. Bartlett in 1932—is certainly relevant to the theory of scripts. The fundamental assumption of script (and schema) theory is that situations can be understood only in terms of the scripts available to the comprehender. Recall the generic restaurant represented in Figure 4.2. It is composed of four different “scenes”: entering, ordering, eating, and exiting. We would not be surprised to find that an individual whose experience of restaurants is limited or non-existent—for example, a member of a hunter-gatherer society or a nomadic Sahara-dwelling Bedouin—has difficulty understanding what is going on in a conventional restaurant episode. Accordingly, one prediction of script theory is that those aspects of stories or situations that do not neatly fit a person’s available scripts will likely be changed and stored in memory in a manner that achieves a better fit. The distortion can occur in two basic ways: new information may be added to make the story more congruent with what the subject knows, or certain information may be deleted to avoid conflict.

Evidence for these predictions is readily found in the literature: Bartlett (1932) reports that stories are altered exactly as expected to fit the conceptualizations of the subject; Bransford and Johnson (1973) report that in an ambiguous story in which a particular sentence is interpretable only under one reading, subjects remember that sentence much better when it is consistent with their reading of the remainder of the story than when it is not; and Spiro (1977) has shown that, after long delays, subjects distort stories with surprise endings to make the ending more probable, where the distortions involved both the addition of congruent elements and the deletion of incongruent ones.
This last finding is hardly surprising: insofar as scripts constitute a theory of how humans store information in memory, it makes perfectly good sense that after long time intervals—when most of the details of an episode have eroded or faded away—recalling the relevant events will be more like a reconstruction process (which naturally draws upon the available scripts in LTM), than a faithful recollection of actual experiences.

What is interesting is that script theory makes that opposite prediction for short time intervals, based on the script (or schema)-copy-plus-tag (SC+T) hypothesis mentioned in §4.3.ii. Evidence for this prediction has been reported by a variety of researchers—including, Bower, Black, and Turner (1979); Graesser, Gordon, and Sawyer (1979); Smith and Graesser (1981); Graesser and Nakamura (1982); Nakamura, Graesser, Zimmerman, and Riha (1985); and Maki (1989). The paper by Smith and Graesser is a good representative of such data.

Smith and Graesser were investigating the role of typicality or relevance of specific events and actions in people’s memory for script-based passages. Do we remember the predictable events and actions better than the unpredictable ones, or is it the reverse? They presented a total of eight passages to their subjects, each one describing a different scripted activity—“Getting Up in the Morning,” “Taking the Dog to the Vet for Shots,” “Washing a Car,” “Cleaning an Apartment,” “Eating at a Restaurant,” “Visiting Someone at a Hospital,” “Going to the Beach,” and “Washing Clothes at a Laundromat”—and tested them with either a recall or recognition task. Tests were conducted thirty minutes after hearing the passages, then again after two days, one week, and three weeks.

The researchers took great care in constructing their passages. Stories mentioned both typical and atypical events within each script situation. Before testing subjects, the investigators collected norms in order to know what is typical and what is not. In their standard analysis of recall and recognition performance, typical information was remembered better than atypical information. These scores were then corrected for guessing, since the high accuracy on typical information probably included both events that were genuinely remembered as well as events that were merely reconstructed from the scripted knowledge. When the scores were corrected for reconstructed guesses, recall and recognition were higher for atypical events than for typical events. For instance, in a story about taking the dog to the vet, subjects showed more accurate memory for the unusual events that occurred (e.g., “while waiting for the vet, Jack dropped his car keys”). Typical events, those anticipated by the script (e.g., “Jack led the dog into the waiting room”), were recalled more poorly. But these results reversed as time went on: eventually, the typical events were recalled much better than the atypical events. As Smith and Graesser summarize (1981: 555),

there was better recall for atypical actions after a short retention interval, but a better recall of typical actions sometime around the 2-day retention interval and thereafter. The crossover is consistent with the notion that schemata become progressively more important in guiding conceptually driven retrieval as the retention interval increases.

These findings suggest that people represent a scripted sequence of events in memory by pointing to a copy of the generic script plus “tagged on” time-sensitive representations of the unconventional features of the experience. So when we hear a script-based story, the script leads us

---

22 A recall task is a memory test in which subjects must reproduce, as accurately as possible, a previously presented stimulus—usually a list of words or a short story. Recall tests may require subjects to repeat items in any order (free recall), or in the same order in which the items were presented (serial recall). A recognition task is a memory test in which subjects are asked whether or not they have seen a particular stimulus before, to which they must give a “yes” or “no” response; Ashcraft (1994).
to expect certain events—the typical events—and provides default values for them. Such default values do not need to be stored in a memory trace, since they already are stored as a part of the script. On the other hand, the script does not prepare us for unusual events (like dropping the car keys). Thus when it is time to recall the story shortly after encoding had taken place, the atypical events benefit from the advantage of having been specifically tagged and stored in memory as different from those actions anticipated by the script.

The overall pattern of recall ... is consistent with the [SC+T] model's assumption that the generic schema becomes progressively more important in guiding conceptually driven retrieval over time. At short retention intervals, memory is primarily reproductive, with most of the information being retrieved from the memory trace. As the retention interval increases, the tagged atypical actions in the trace become less accessible because retrieval becomes more dependent upon the generic schema. As retrieval becomes more dependent upon the generic schema, memory becomes more reconstructive. (Smith and Graesser 1981: 557)

Nakamura, Graesser, Zimmerman, and Riha (1985) set out to test the Smith-Graesser results in a more naturalistic setting, using a classroom lecture as the input and a later memory test as the evidence. As was found with the prose passages, memory was better for the atypical or irrelevant information (e.g., sipping a cup of coffee) than for more typical information (e.g., underling a word on the blackboard). Part of the strength of the Nakamura et al. paper is that it extends the script approach to settings involving more than written or spoken passages. Another strength is that the results were obtained in a "blind" procedure: the students were unaware that their memory would be tested for events that happened during the lecture.

The script-copy-plus-tag hypothesis takes for granted the idea of default value, which tells us that not everything need be mentioned in a story for comprehension to take place: "When someone decides to tell a story that references a script, he recognizes that he need not (and because he would otherwise be considered boring, should not) mention every detail of his story. He can safely assume that the listener is familiar with the referenced script and will understand the story as long as certain crucial items are mentioned" (Schank and Abelson 1977: 38). Accordingly, script theory predicts that people's recall of a story or a workaday experience will be influenced not merely by details that were mentioned or observed, but also by the events and details that were inferred based on scripted knowledge. As a simple example, in a restaurant story one might "recall" that a customer left a tip for the waiter even though no tip was ever mentioned in the passage. Where does this information come from?

To answer this question, Bower, Black, and Turner launched an immense study, consisting of seven separate experiments, involving hundreds of subjects:

Our experiments investigate some psychological implications of Schank and Abelson's script theory. Experiments 1 and 2 examine the organization of people's knowledge about stereotyped activities. What actions, roles, and props do people mention and how do they group or cluster these into subscenes? Experiments 3 and 4 ask whether, in remembering a text mentioning a subset of script actions, people tend to remember numerous unmentioned parts of the underlying script. Experiment 5 examines whether in recalling a text people will tend to recall the script actions in their stereotypic order even though the actions are mentioned in another order in the text. Experiment 6 asks whether the reading of earlier actions in a script speeds up the reading and comprehension of later actions in that script. Finally, Experiment 7 examines memory for occasional events, inserted into script-based stories, which interfered with or deviated from the smooth-running of the script. (Bower et al. 1979: 180)

---

23 As does the work of Pezdek et al. (1989), in which the same effects were reported in a variety of real-world scenes.
The result of the first experiment, in which people described in detail what goes on during familiar activities, found that subjects generally agreed on the nature of the characters, props, actions, and the order of events in a script. "What is surprising is how much agreement there is in the ‘basic action’ language that people used to describe the activities. This uniformity is reflected in how few of the events were mentioned by only one person. For example, in the restaurant script, of 730 actions mentioned in total ..., only four were completely unique (given by a single person)” (1979: 181).

Scripts are not undifferentiated, linear chains of occurrences, but seem to be organized into major “scenes,” with those composed of sub-sequences of actions. Script theory therefore predicts that people should coincide on the demarcation of the various segments of a generic sequence of events. Bower et al. sought to test that hypothesis in experiment 2, and found that “subjects agreed ... that a continuous script activity can be segmented into chunks or scenes. And they agreed with one another where the scene boundaries were located in the event sequences” (1979: 186). In one case, the agreement reached 99%.

Next, Bower and his colleagues investigated recall of texts composed of selected lines from an underlying script. The question was whether in recalling such a text subjects will use the underlying script to “fill in” gaps of intervening actions not explicitly mentioned in the text. They found that (i) subjects did indeed “recall” gap-filling events left out of a script story, and did so more if they were exposed to other script instantiations that did mention the analogous actions (experiment 3); (ii) that there was a bias resulting in a higher rate of false recognitions for typical actions that were not mentioned in the text (experiment 4); (iii) that subjects preferred to recall scripted actions in their “canonical” order (experiment 5); and (iv) that the reading time (which is normally taken to be a measure of comprehension) for a particular passage increased to the extent that the events described in the passage were unanticipated by the underlying script (experiment 6).

Finally in experiment 7, Bower et al. found that for short time intervals (three days at most), subjects remembered interruptions in the flow of a script “better than the routine script actions because [the interruptions] will appear subjectively more important and so will occasion more attention or deeper processing,” (1979: 210) just as predicted by the SC+T hypothesis.

In a similar connection, several studies have reported that bridging inferences, or coherence-based inferences, are generated during comprehension (Long et al. 1990; Bransford et al. 1973). Bridging inferences serve to fill conceptual gaps in the explicit text of a story. They are needed to establish a coherent representation of events. There are several types of bridging inferences: (a) inferences specifying that the same words across two propositions refer to the same concept or script (McKoon and Ratcliff 1980); (b) inferences specifying that an anaphor and its antecedent refer to the same concept or script (McKoon and Ratcliff 1980); and (c) inferences which establish the causal connections between the events in the story (Graesser and Clark 1985). Bransford and Johnson (1973) report evidence for these and other categories—including, spatial relations among objects, instruments used to carry out tasks, and antecedent conditions relating two or more events. Causal inferences are especially interesting, since they reveal more than the others the kind of world knowledge that comes into play during the process of comprehension. Graesser and Clark (1985)

---

24 This hypothesis is endorsed by the experimental work of Abbott, Black, and Smith, who found that “actions in scripts are linked together in memory as sets; that is, when some of the actions in the set are accessed, so are the others” (1985: 181).

25 The same result was also found by Graesser and his colleagues (1979), and Seifert and her colleagues (1985).

26 An “anaphor” is a word or phrase that takes its reference from another (preceeding) word or phrase.
have undertaken an investigation with the aim of identifying the types of knowledge-based inferences that are generated in response to short narrative texts such as *The Czar and His Daughters*:

Once there was a Czar who had three lovely daughters. One day the three daughters went walking in the woods. They were enjoying themselves so much that they forgot the time and stayed too long. A dragon kidnapped the three daughters. As they were being dragged off, they cried for help. Three heroes heard the cries and set off to rescue the daughters. The heroes came and fought the dragon and rescued the maidens. Then the heroes returned the daughters to their palace. When the Czar heard of the rescue, he rewarded the heroes. (Graesser and Clark 1985:86)

Graesser and Clark have adopted a question-answer method for exploring the generation of causal inferences. As subjects read a text—incrementally, sentence-by-sentence—they answered questions about each of the explicit statements. Answers to questions such as “Why?”, “How?”, and “What happened next?” have uncovered a large corpus of script-based inferences that readers generate as they process a narrative. Moreover, by manipulating various contextual elements in the stories, Graesser and Clark were able to observe which explicit statement or group of statements activated each of the script-based inferences, and which knowledge structures contribute to the process of inference generation. They then developed a system—a conceptual graph structure—for mapping out the relevant world knowledge which supplies the inferences in question.

A small portion of the conceptual graph structure for the *Czar and his Daughters* is depicted in Figure 5.3 (adapted from Graesser and Clark 1985). The statement nodes in squares refer to explicit statements in the text. The statement nodes in circles and ovals are inferences generated by the question-answer procedure. The directed arcs correspond to the arc categories reason (R), consequence (C), initiate (I), and outcome (O).

Two points. First, the research by Graesser and Clark shows just how much people draw upon background knowledge during the understanding process, which seems to involve the capacity for both explicit and implicit inference-making. Why would people infer that the dragon kidnapped the daughters in order to kill and eat them, unless people have a DRAGONDS which contains the information that dragons are capable of such behaviour? And why would readers infer that the Czar’s daughters were frightened by the kidnapping, unless some generic kidnapping scenario (a KIDNAPes) includes the information that kidnapping causes fear on the part of the victim(s)? I take these conclusions to follow from a fairly basic truism: intelligent as we may be, humans have still not acquired the ability of creating information from thin air. If it isn’t represented in the brain, it isn’t available!

Second, aside from obvious notational differences, Graesser and Clark’s conceptual-graph model is congruent with Schank and Abelson’s CD theory: both use more or less the same categories and primitive acts. In fact, there is no reason to believe that the representation of *The Czar and His Daughters* in Figure 5.3 cannot be “directly translated” into Conceptual Dependency. Since conceptual graphs and CD were developed independently, each lends support to the other, and suggests that script theory—being the offspring of CD—rests on a good foundation.

Most studies reported so far have a common property: the acquisition materials have involved written passages. What happens when stimuli are non-verbal? Graesser and Nakamura (1982) compared memory for actions in videotaped scripted activities and audiotape-recorded

---

27 For further details regarding this method, see Graesser, Robertson, Lovelace, and Swinehart (1980).
activities. The action sequences in the videotape paralleled exactly the sequences described in the audio recording. Four experimental scripts were used: "setting the table," "polishing shoes," "fixing lunch," and "typing a letter." The tape-recorded versions were read at a medium rate of approximately 150 words per minute, and all videotaped episodes were presented in about ten minutes each. Thirty subjects were assigned to the videotape condition and thirty subjects were assigned to the audiotape condition. Approximately fifteen minutes after viewing or listening to the scripts, subjects completed a recognition test. No significant differences between the two conditions were found: both groups showed the same bias toward atypical events in the short-term and a bias toward typical events in the long-term (once more supporting the SC+T hypothesis); the long-term reconstruction of event sequences tended toward the underlying generic scripts for all subjects; both groups made the same bridging inferences and had similar false-alarm rates, and both reconstruct the sequences into familiar "scene" subdivisions. These findings have prompted Graesser and Nakamura to conclude that scripts are robust entities that operate over multiple knowledge domains.  

There is an interesting line of research that supports the Graesser-Nakamura conclusion—suggesting that scripts also operate in visual information processing. Pezdek and Chen (1982) investigated this possibility by presenting to subjects line drawings of scenes that were either simple (without much elaborative detail) or complex (with some elaborative detail). The pictures were later tested in either the same state or in an altered state—with simple pictures changed to their complex

---

28 Minsky (1975) has done a good deal of work in this area.
form, and complex pictures changed to their simple form. The subjects' task was to indicate whether the test pictures were the "same" or "different" (see Figure 5.4).

Figure 5.4: Examples of pictures in their simple and complex forms. The two versions of each picture contain the same central information, but extra peripheral details, shading and embellishment were added in the complex pictures. The details were added with the proviso that they be irrelevant to the main theme of the picture. In some cases the extra details were added primarily to the figure itself, while in others they were added to the background. Pezdek and Chen do not believe that the difference matters (Adapted from Pezdic and Chen 1982).

Pezdek and Chen found that changes from simple to complex (i.e., addition of details) were recognized better than changes from complex to simple (i.e., deletion of details). They hypothesized that pictures are schematically encoded in such a way that the memory representation of both simple and complex pictures is similar to the simple version of each picture. Changes involving added details were recognized as being different from the simple pictures originally presented because the added details were never part of the visual script in the first place. On the other hand, simple pictures that were originally complex are not recognized as different because, despite the deletions, the simplified versions still match the visual scripts used to process the images. Pezdek and Chen call this the asymmetric confusability effect.

The asymmetric confusability effect was later verified by Pezdek et al. in 1988. It has also been observed by Agostinelli, Sherman, Fazio, and Hearst, who reported that subjects were more accurate at detecting and identifying additions than deletions to line drawings of objects. And Healy (1981) showed that in the process of proofreading, subjects are less likely to notice missing features of a letter (e.g., when "students" is misspelled as "students") then they are to notice added features of a letter (e.g., when "factors" is misspelled as "faetors"). The asymmetric confusability effect thus appears to be a general feature of schematic processing.

According to which people have more difficulty in perceiving changes in complex rather than simple patterns because the proportion of change for a stimulus with more information is less than the proportion of change for a stimulus with less information. To show that the asymmetric confusability effect is not an instance of Weber's Law, Pezdek et al. (1988) ran the "changed picture" test again, this time with a one-sentence description of each picture included in the presentation. What they found was that the sentence exaggerated the asymmetric confusability effect, a result for which Weber's Law gives no
In an effort to determine whether Pezdek's model generalizes to semantic memory, Bharucha et al. (1985) investigated subjects' abilities to detect changes in prose. The subjects were presented with pairs of sentences, some of which were "coherent," such as

\[(1)\] The ship pitched wildly in the storm. 
John hoped the weather would clear up.

For each coherent pair, a corresponding "anomalous" pair was constructed by replacing a target word so as to change the meaning of one of the sentences. Hence, replacing "weather" with "acne" yields the following anomalous pair:

\[ (1^*) \] The ship pitched wildly in the storm. 
John hoped the acne would clear up.

The test phase required subjects to indicate whether each pair of sentences was the "same" or "changed" from its original form. Changes were either from coherent to anomalous or from anomalous to coherent.

The researchers reasoned that during the study phase, coherent sentence pairs like (1) will activate in subjects a unifying cognitive unit representing, e.g., a stormy sea voyage; sentence pair (1*) will activate two abstract cognitive units, one representing a stormy sea voyage and the other an acne problem, with no obvious link between them. If (1) occurs in the study phase and (1*) in the test phase, the second test sentence will not be assimilated to the stormy sea voyage and will easily be detected as new. If (1*) occurs in the study phase and (1) in the test phase, the second test sentence will be assimilated to the stormy sea voyage, making a false alarm likely.

Results showed that subjects were in fact more likely to notice changes from coherent to anomalous than the other way around. Bharucha et al.'s account of this phenomenon is essentially the same as Pezdek et al.'s explanation of the same result in regard to visual memory: because sentences activate specific schemata, changes that can be incorporated within these schemata are not easily detected, but changes that cannot be so incorporated tend to stand out.

The studies discussed in this chapter are only a tiny fraction of what can be found in the literature. The corroborative evidence is enormous. Many of the above experiments were duplicated, time and time again, with the same findings. The theory of scripts is the glue that binds all these results together. At least some human Intentional states are script-based, and to speak about the Intentional content of these states is to speak about the content of these scripts. We have not discussed content in this chapter, but it is a topic that we shall not delay any further.

---

31 The asymmetric confusability effect has also been observed in musical materials. An alteration of a tonal melody is more accurately detected if the tonal structure is violated than if it is preserved (Cuddy, Cohen, and Miller 1979). The same is true of sequences of chords (Krumhansl, Bharucha, and Castellano 1982). These parallels point to what should by now be clear: principles of cognitive organization (scripts) cut across propositional and non-propositional domains.
Also, there are a number of things that people commonly profess to know ... for which there is no definite piece of evidence, no single state of affairs or easily specifiable set of such states, that even approximates a conclusive reason ... Countless experiences converge, so to speak, on the truth of a given proposition ... The fallibility of source A and the fallibility of source B does not automatically entail that when A and B agree about P's being the case, that ... P might still be false.


In the beginning there was information. The word came later.


I find myself sympathetic to what Andy Clark calls historical realism. "The historical realist holds (with some refinements) that belief talk exhibits the contents to which an organism became sensitized by learning during the course of its cognitive development. How the learning proceeded (what kind of learning algorithms were used) and how the resulting sensitivity is realized (what kind of storage and recall mechanisms are available) is irrelevant. All that matters is that the learning took place, and that it set up links to the various external states of affairs to be cited in basic belief talk" (Clark 1990: 96).

What I like about this position is that it suggests cogent limits on the obligations of philosophers of mind (limits, incidentally, which are already observed in practice even if contestable in ideal). To put it bluntly, the philosopher’s job ought not to include experimental investigation, since there are others who are more qualified to evaluate the veracity of the various cognitive theories on offer and decide which among them best fits the empirical facts. The suggestion is for a collaborative division of labour between the philosopher of mind and the natural scientist. Let the former pursue theory construction—not haphazardly, but by taking account of the relevant findings of the natural sciences—without worrying overmuch about the tedium of (but nevertheless still yielding to) the corroborative process; and let the latter’s job be that of empirical testing.

Clark admits that his expression of historical realism leaves room for “refinements.” For human organisms, the improvements I would make are, first, that “propositional attitudes” be substituted for “belief” throughout the passage. I think the substitution is proper, because the contents which arise in answer to the learning and the subsequent sensitization of a person during his or her cognitive development need not enter only in beliefs, but may just as well be implicated in desires, fears, hopes, and other psychological states. That is to say that organisms capable of
Intentional thought may come, not only to believe that such-and-such as a result of learning, but also to desire, hope, or fear that such-and-such as a result of learning.¹

Secondly, though the specifics about how the physical realization of an organism’s learning-induced “sensitivity” may ultimately be irrelevant to some philosophical questions, it is not irrelevant to all—e.g., those surrounding connectionism and computational psychology—and may therefore benefit from some theorizing. Again, such theorizing cannot be carried out fancifully, but must be shaped by the relevant scientific discoveries.

I shall do my best to follow my own advice in this chapter, which is concerned with the physical implementation of scripts. I admit at the outset that nothing definitive will be offered. What I present is an amalgam of different suggestions, clues, conjectures, and hints that I stitch together in order to fashion one plausible way of accounting for scriptal brain-level functioning. What will emerge from this patchwork is a sketch of what makes human scriptal structures contentful. But I am neither interested in providing a comprehensive picture (because I can’t), nor in overthrowing alternative interpretations (because I haven’t the means). And although I take it as a merit that what I have to say is susceptible to empirical testing, as I lack the expertise to assess the veracity of my proposals with experiments, I must leave the verification work to others.

Our first stop is Dretske’s information-processing account of Intentionality and mental causal efficacy.

### 6.1 The Informational Origins of Thought

Fred Dretske has combined the best of Fodorian representationalism with the teleological approach of Millikan and McGinn. Dretske’s first major work was *Seeing and Knowing*, published in 1969. In a sense, the title of this book sums up its author’s long-term project of showing how, by beginning with a description of those input mechanisms we call the senses, and by treating the brain as a kind of data-driven engine, we can construct a reasonable account of complex mental states and cognitive faculties.

The first part of Dretske’s program was taken up in *Knowledge and the Flow of Information*.² For Dretske, information is a commodity that must produce knowledge in a recipient who is properly attuned to the vehicle of that information. If I utter falsehoods, though the sentences I use may be perfectly grammatical and convey all their conventional meanings, I impart no information. It is this semantic aspect of information—the fact that it embodies content—that is central to Dretske project of naturalizing the mind. To that end, he makes a number of important distinctions.

One is the distinction between information as carried and information as received. Information at the source of a communication channel is much more diffuse and scattered than the same information after it has reached a recipient. This is because a recipient often has a lot of background knowledge which acts as a “sieve” on incoming data. Lyons (1995: 98) gives the following analogy:

---

¹ Hence Searle’s (1979) distinction between the *representational content* of a psychological state (e.g., that Delilah has a pet iguana) and the *mode* of that psychological state (e.g., believing, hoping, desiring, or fearing that Delilah has a pet iguana).

² Alston (1983), Loewer (1982), Maloney (1985), and Sterelny (1983) are just a few of the many reviewers of this book.
If I were, say, to play a game of ‘Twenty Questions’ in order to discover what object it is in this room that you are thinking about, I could begin with the question ‘Is it in the left-hand side of the room or the right-hand side?’ Your answer to my question will immediately halve the possibilities. Then if I ask the question ‘Given that you have said that it is in the left-hand side of the room, is it in the northern half of that half or not?’ Your answer will again halve the possibilities. And so on. Each answer to my question increases my information, and this increase can be quantified as a ratio between the possibilities existing before any answer is given (or the ‘amount of information at the source’) and the possibilities existing after the last answer is given (or the ‘amount of information at the receiving end of the communication chain’).

Even when information has been received, it still remains unusable until it has been selectively sifted by the human perceptual and processing systems. This initial diffuseness of data is not just a matter of its sheer volume; it is also a matter of its (temporary) lack of organization. The kind of organization required will, of course, depend on the type of information-processing system we are dealing with. In the case of human beings, a good deal of data must be organized conceptually—i.e., in terms of concepts and their relations—since this is how people understand the world.

Dretske also makes use of the distinction between digital and analogue encoding. A digital device is a binary device, one with only two discrete informationally relevant states. Thus a simple “light-on/light-off” electrical lamp digitally encodes the information that either current is flowing through the system or that it is not. There are no informationally relevant intermediate states. On the other hand, a light bulb that has many positions—that is, one that can gradually brighten up from a position of emitting no light at all to one of emitting light of such-and-such candlepower—is an example of an analogue system. The on-off lighting system will indicate quite definitely the informational state it is in, because there are not many states to choose from: either the current is flowing or it isn’t. The variable-brightness system, however, will have many more possible positions, but much less precision about its informational state.

Dretske (1981) puts forward the view that human perception is produced by analogue systems whose information becomes transformed into cognitive content precisely insofar as the information becomes digitally encoded. Humans are hybrid digital-analogue processors, but the interplay of our perceptual and cognitive systems is based on the transformation of data from analogue to digital form. Indeed, it is the “successful conversion of information into (appropriate) digital form that constitutes the essence of cognitive activity” (1981: 142). So, for example, my olfactory sense might register in an analogue way the presence of some odour in the kitchen that could be Gorgonzola cheese, aged Prosciutto ham, or yesterday’s Chicken McNuggets; but I will not know what is in the kitchen, not even this disjunctive list of possibilities, until some information is extracted and encoded digitally. Another way of putting this is to say that the transformation from sensory data to real knowledge involves the generation of specific contents which produce beliefs, and “to occupy a belief state a system must somehow discriminate among the various species of information embodied in a physical structure and select one of these specific pieces for further treatment—as the content of that higher-order Intentional state that is to be identified as the belief” (1981: 174).

---

3 W. P. Alston seems to miss this point, writing that “what sensation carries ... is too bulky to be used by the subject, all in a piece, for the guidance of behavior, for inferential extension of knowledge, and so on. So the distinction really has to do with whether the total information carried by a state is usable as such by the organism” (1983: 453-4).

4 This, by the way, is how Dretske grounds his distinction between “epistemic” and “non-epistemic” perception in Seeing and Knowing.
But how can physical structures that carry data in analogue form be transformed into physical structures that carry data in digital form? Dretske tells us that the meaning of a structure is derived from its informational origins. It is here that his account connects with evolution, species adaptation, genetic inheritance, and especially individual learning and maturity.

That is to say, a certain type of structure acquires its content, the sort of content we associate with belief, by its informational origin. An internal structure develops (during learning) as a system's way of completely digitalizing information about, say, the F-ness of things. This way of encoding information (as the semantic content of an output-determining structure) makes the information so encoded relevant to explaining the system's behavior. It is this origin that defines the content or meaning of the internal structures. It defines what a system believes when one of these structures is later instantiated with respect to some perceptual object. That a system believes something depends, partially, on the effects (on some output) of these internal states, since to qualify for cognitive content an internal structure must have executive responsibilities. But the content is determined solely by the structure's origin—by its information heritage. (1981: 201-2)

In *Explaining Behavior*, Dretske turns to the second part of his project—i.e., providing a theory of how beliefs and desires are to be identified with internal information-carrying representations in the brain that cause bodily movement (or an inhibition of movement) in virtue of the information they carry. This part of Dretske's project adds considerably to the work already done in *Knowledge and the Flow of Information*. In particular, it explains the brain's information-bearing mechanisms and their functional role by means of a sophisticated account of representational systems and reason-guided movements.6

To get genuine meaning for the "indicator circuits" in an organism's brain, we need to turn to cases of individual learning in which "internal states acquire control duties or change their effect on motor output as a result of their relation to the circumstances on which the success of this output depends" (1988a: 95). As an example, Dretske cites the kind of learning exhibited by some parallel distributed processing (PDP) networks. What is significant here is that internal structures (weights on connections) acquire their causal roles in response to the consequences of the weightings. If the consequences are positive—i.e., if the network gets the output right—the weightings are retained; if not, they are automatically adjusted. In this way internal structures are recruited to support a desired output. Likewise, a pigeon that learns to peck a target when a light comes on may be said to have recruited an internal state to guide its output (pecking) because of what the state indicates (the presence of food in a box, say). "Since these indicators are recruited for control duties because of the information they supply, supplying this information becomes part of their job description—part of what they, once recruited, are supposed to do" (1988a: 99).

Dretske thus defines a representation system as one whose function is to indicate something about an object, or condition, or state of affairs outside itself. He then goes on to distinguish three kinds of representational systems. First we have conventional signs, or Type I representation systems. A good example of these is the "♀" symbol, which is used to denote the female gender. The

---

5 See the review by Meyers (1989) and Clark's (1990) critical study.

6 The central arguments of *Explaining Behaviour* are summarized nicely by Dretske in a short paper called "The Explanatory Role of Content." In this paper, Dretske makes it clear that he is "a realist about content, and my realism stems from a conviction that content has an essential and ineliminable role to play in the explanation of behavior. We don't describe one another as believing this or desiring that because it is, as it is with certain machines, merely a usable predictive strategy. Nor do we do it because we are too ignorant of neurobiology to know what is really in there making the limbs move. We describe each other in this way because our inner states actually have a content and it is this content that explains why we do what we do" (1988b: 32).
meaning of "♀" is purely arbitrary; its representational function is conferred on it by humans and could have been something entirely different.  

Second, there are natural signs—or Type II representation systems—such as footprints, cloud patterns, and polygraph tests. These structures or formations become signs insofar as humans, recognizing their natural causal connections to other objects or states, use them as indicators of those objects or states. What is natural (and so intrinsic) to such signs is that they form part of some causal network. What is conventional about these signs is that we humans, having worked out their causal relations, adopt the cause (or part of the cause) as a sign of the effect (or part of the effect).

Finally we come to wholly intrinsic natural indicators. "Natural systems of representation, systems of Type III, are ones which have their own intrinsic indicator functions, functions that derive from the way the indicators are developed and used by the system of which they are a part. In contrast with systems of Type I and II, these functions are not assigned. They do not depend on the way others may use or regard the indicator elements" (1988a: 62). Dretske points out that every animal depends on wholly intrinsic natural indicators for its survival. All organisms have sensory systems which help to construct internal structures that carry information about their environment, thereby functioning as behaviour-guiding devices. These internal structures are created in part by the selectional pressures of the environment in the history of the species of which that individual animal is a member, and in part by the influence of the environment during the local history of that individual.

As Intrinsic natural indicators, the contents of information-registering states of Type III systems have reference as well as sense. Reference is what the indicator is about, what it is in the environment it picks out; sense is the particular form the indicator takes, the way in which the referenced object is represented. Like natural signs, intrinsic natural indicators are constructed entirely by the causal powers of nature. But unlike natural signs, their deployment as action-guiding mechanisms is internal to the system. So Type III representations can be said to have original Intentionality. Furthermore, the very same internal structure can act both cognitively and conatively, can be both a belief and a desire; for these roles are functional ones. It is the use to which these internal representational formations are put that makes them one or another attitude.

The reason we attribute the titles "beliefs" and "desires" to the internal representational structures of humans but not to "lower" animals or artefacts is partly a matter of convention. The other part is related to the fact that, in humans, these internal representations have a holist character: they are intertwined and interconnected with numerous other such structures based on their semantic properties. Moreover, human internal information-bearing mechanisms are modifiable by learning; i.e., they are plastic. While other animal species exhibit signs of plasticity, none comes close to humans. For people, the manner in which a goal is accomplished is often cognitively fluid and behaviourally underdetermined, but in the case of many (probably most) non-human animals, underdetermination is often exhibited only at the very basic level of physical movement.

---

7 It could still be different. All we need to do is change the relevant convention(s).

8 This is where the kinship to Goodman’s distinction between o-pictures and pictures-of-o is most manifest in Dretske’s theorizing; see §2.2.i.

9 There is no tension between this claim and Dennett’s observations regarding the prevalence of the Intentional stance. Dretske is not denying that we do seem, at least in common discourse, to take up the Intentional stance toward non-human animals and sometimes even toward machines. Rather, he believes that we would be reluctant to do so if entreated to speak literally rather than metaphorically, and I think most of us probably would.
Dretske limits the term “behaviour” to bodily movements that are the outcome of certain chains of events. These events begin with intrinsic natural indicators, which cause our muscles to engage and so move our bodies in the service of higher-level goals. Behaviour is thus a process and not just a final outcome. Confusing behaviour with output leads to the myth that we can describe what we do by offering neurobiological explanations of movement, when we should instead be offering psychological explanations of actions. Accounts based on reasons are therefore not in competition with neurophysiological reports: the latter explain how an action came about, the former explain why an action occurred.10

Finally, in *Naturalizing the Mind* Dretske expands the teleological aspects of his program to cover the qualitative or what-it-is-like characteristics of sense experience, introspection, qualia, and consciousness:

A state is a conscience experience of $F$ (thus making its possessor sensually aware of $F$) if the state has the natural (systemic) function of providing information about the $F$-ness of objects standing in the appropriate contextual relation ($C$) to the system. States thus become conscious by acquiring, for some determinable property $F$, the function of indicating, for appropriately related objects, their determinate value of $F$. Since states acquire their systemic functions through evolutionary processes—here assumed to be natural selection—natural selection is being identified as the source and creator of conscious experience. (1995:162)

Thus Dretske makes no secret of the externalist dimensions of his views. The version of RTM he sanctions is one that identifies mental facts with representational facts. But, though representations are in the head, the facts that make them representations—and, therefore, the facts that make them mental—are outside the head. A state of the brain is an experience only if it represents the world in a certain way, and a state represents the world in this way only if it has an appropriate information-carrying function. Since functions have to do with the history of the states and systems having these functions, Dretske believes that mental facts do not supervene on what is in the head. What is in heads $A$ and $B$ could be physically indistinguishable and yet, because their owners have had relevantly different histories, one is a representational system, the other is not; one makes the person in whom the states occur aware of the world, the other does not. Or, in any case, so says Dretske.

6.2 Naturalizing Scripts

I think we can find in Dretske much of the raw material needed to create a blueprint of how the script construct might be realized at the neural level of the brain. We have talked about four different kinds of scripts—episodic, instrumental, personal, and definitional—but starting now, scripts will gradually dominate our discussion, since they are especially important to the upcoming philosophical issues surrounding Intentionality and narrow content. Nevertheless, everything said about the physical basis of scripts is meant also to apply to the other script types.

---

10 In a similar connection, Lyons rebukes what he calls *philosophical anorexia*: “Choosing the right level of explanation is not necessarily non-reductionist. It may involve what might be called non-anorexic reduction, that is to say, a reduction down to a lower level which proves to be the right one together with a principled refusal to go further down, precisely because the right level has already been found … What the recent history of philosophy of mind has taught us to eschew are those reductions which try to produce an explanation at such a low level as to be useless and so inappropriate. It has taught us to avoid philosophical anorexia … [T]his tendency to philosophical anorexia has often arisen among philosophers of mind through their incontinent adulation of the oh-so-glamorous and oh-so-slim theories that strut along the catwalk of the hard, experimental, physical sciences” (2001: 218-20).
i. Object Recognition and Simple Definitional Scripts

It is not just a naturalization of scripts that we want to achieve. We also want to explain how scripts get their content. The key idea in both cases is the transformation of data from analogue to digital form. This was a central notion for Dretske, who took a good deal of time and care in presenting it:

To illustrate the way this distinction applies, consider the difference between a picture and a statement. Suppose a cup has coffee in it, and we want to communicate this piece of information. If I simply tell you, “The cup has coffee in it,” this (acoustic) signal carries the information that the cup has coffee in it in digital form. No more specific information is supplied about the cup (or the coffee) than that there is some coffee in the cup. You are not told how much coffee there is in the cup, how large the cup is, how dark the coffee is, what the shape and orientation of the cup are, and so on. If, on the other hand, I photograph the scene and show you the picture, the information that the cup has coffee in it is conveyed in analogue form. The picture tells you that there is some coffee in the cup by telling you, roughly, how much coffee is in the cup, the shape, size, and color of the cup, and so on. (1981:137).

Consider another simple illustration. A variable source (the speed of a car) is capable of assuming 100 different positions (from 0 to 99 mph). Data about this source is fed into an information-processing system, the first stage of which contains a device (a speedometer) that accurately registers the state of the source. The information from the speedometer is then fed into a converter which consists of four differently pitched tones: if the source is in the 0-14 mph range, the lowest-pitched tone is sounded; a higher-pitched tone is sounded for the range 15-24 mph, a still higher pitch from 25-49 mph, and the highest at 50-99 mph. “These different ranges may be thought of as the approximate ranges in which one should be in first, second, third, and fourth gears, and the converter a device for alerting novice drivers (by the differently pitched tones) of the need to shift gears” (Dretske 1981: 140).

Notice that the output of this system always carries less information than the input: whereas input carries information about one of 100 different states of the source, the output carries information about one of only four states. What is gained by this loss of information is a classification (in the various tones) of the significant ranges of the input variable. This is a (primitive) form of stimulus generalization: the information-processing system ignores the difference between 32 mph and 43 mph, treating both values as the same because both activate the third tone.

From the point of view of the information the system is designed to communicate, the internal speedometer is an analogue representation of the state of the source, because it carries much more specific information than is required to control the system’s output. Suppose the speedometer indicates that the state of the source is 32 mph. Nested in this piece of information is the further datum that the source is travelling between 25 and 49 mph. The digital converter, comprising the four differently pitched tones, is “interested” only in the latter fact. Hence, as far as output is concerned, the information-processing system does not discriminate between 32 and 43 mph. In fact, this is the whole point of the four-tone converter: to prevent the system from discriminating between these (and other) states of the source.

Ultimately, then, a process in which information is converted from analogue to digital form is a process that involves the loss of information. Data is lost because we pass from a structure of greater informational content (the speedometer) to one of lesser informational content (the four-tone converter), where irrelevant details (the precise speed of the source) are pruned away and discarded. Until information has been filtered in this way, Dretske assures us that an information-processing
system has failed to treat different things as essentially the same. It has failed to classify or categorize or generalize, failed to recognize the input as an instance (token) of a more general type.

Perception is one such process. It is a process by means of which data is delivered within a richer matrix of information to the cognitive centres for selective use. Seeing, hearing, smelling, tasting, and touching are different ways of getting information about a source-object to a digital conversion unit whose function is to extract pertinent data from the signal for purposes of identifying the object. If the information that \( o \) is \( F \) is never converted from sensory (analogue) to cognitive (digital) form, the system has perhaps seen, heard, or smelled an \( o \) which is \( F \), but it has not seen, heard, or smelled that \( o \) is \( F \). Cognitive activity, on Dretske’s view, is the conceptual mobilization of incoming information, and this conceptual treatment is fundamentally a matter of going from the concrete to the abstract, of passing from the particular to the general. It is, in short, a matter of making an analogue-to-digital transformation.

What information is isolated from a signal coming from an object depends not only on the physical properties of the object itself, but also on the role the object plays in the episode in which it appears. One typically learns the concept \( F \), learns what an \( F \) is, by being exposed to a variety of episodes, wherein some of the objects are \( F \)s and some are not \( F \)s. Not only is one exposed to \( F \)s and non-\( F \)s, but usually the fact that \( F \)s are \( F \)s and non-\( F \)s are not \( F \)s is evident. In other words, the episode is one in which the information that \( o \) is \( F \) (or not \( F \)) is available to the subject whose discriminatory and identificatory responses are then accordingly shaped.

Dretske’s account of the formation of simple-object concepts meshes nicely with the one given by Schank and Abelson in Scripts, Plans, Goals, and Understanding. Like Dretske, Schank and Abelson believe that the things we see and use in ordinary life are defined through the course of experience.11 So a child learns what a spoon is by generalizing the salient roles spoons have played in past episodes, a process by which spoons are simultaneously assigned functional and physical properties. Participating in events, whether directly or indirectly, is therefore central to the Schank-Abelson theory of cognitive growth, which is partly why these authors express a proclivity toward an episodic organization of LTM.

Merging Dretske’s account of simple-object recognition together with the outline of definitional scripts sketched in §4.4, we get something like this:

I. What is fashioned when one learns to recognize \( o \) as \( F \) is a script\(_D\) of \( o \), and that script\(_D\) is to be understood as an internal network of neural circuits that have acquired, through the episodic experiences of the agent, selective sensitivity to those properties of \( o \) that make \( o \) an \( F \). Thus the process by which analogue information about \( o \) becomes focused on the “\( F \)-ness” of \( o \) is the process by which certain neural structures adapt so as to react discriminately to specific features of \( o \); \( o \)’s definitional script thereby attains the partial function of indicating the presence of the properties to which the internal neural structures constituting \( O_{0s} \) have become responsive, thus enabling the agent to identify \( o \) as \( F \).

I want my motivation for describing scripts in terms of neural structures to be well understood. It should be obvious from our discussion of scripts in chapter IV that scripts come in grades. Some are

11 Of course, this feature is not uniquely shared by the explanations given by Schank-Abelson and Dretske. Many other theories take our concepts to be based on experience. But since my aim is to augment script theory with the account given by Dretske, the similar way in which both deal with the role of experience in cognitive development is worth at least a comment.
simple concepts, like ROCKs; others are very complex, resembling mini-theories—like MY-GRANDMA{s}. It would have been nice if we had a uniform, neat construct which we might call a script, but this is just not so. My concept of a rock—owing to my place in the world, the kind of objects rocks are, the extent to which rocks have impacted my cognitive development, etc.—will not be nearly as intricate as my concept of my grandmother—who is, after all, a living, breathing entity for whom I feel affection and with whom I interact on an equal level. Be that as it may, we still call both ROCKs and MY-GRANDMA{s} “concepts” (hence “scripts”), being careful to properly qualify this designation accordingly as we judge the concept in question to be simple, complex, or something in between.

One sure way to render scripts suspicious is to think of them as abstract structures, as merely a “convenient” or “helpful” way to think about human psychology; for by so doing we also render all of the cognitive abilities which scripts help to explain completely mysterious. A dualist might find solace in this; I don’t. I think of scripts much more concretely, as actual neural structures in the brain. Thinking of scripts in this way raises the question of the level at which they operate. Obviously, scripts function at a level high enough to permit psychological explanations of behaviour—the level of recognition, conscious planning, thinking, and so forth. But such capacities do not come from nowhere. There must be brain-level mechanisms that make even the most complicated cognitive tasks possible. This is a common-sense sort of materialism that resembles the outlook expressed by Galen Strawson in the passage quoted in §1.2.iii.

We can adhere to this materialist attitude while simultaneously heeding Lyons’ warning against philosophical anorexia: we can learn to shun those reductions which try to produce psychological explanations at such a low level as to be useless, without denying that high-level cognitive faculties have neural underpinnings. This is possible because we are not trying to explain anything by claiming that scripts are realized in neural structures; we are simply taking for granted a strictly materialist ontology. Indeed, we have no real alternative. We know that scripts—providing the kind of models to which Craik and McGinn have made reference—do not exist in invisible suspension and simply stick to our heads each time we have need for them. Our concepts of the world around us and the things in it—the simple objects as well as the complex ones—are and must be physically implemented; and this I unapologetically assume to be the case.

On the present proposal, then, scriptsD are the Dretskean neural configurations that arise in consequence of analogue-to-digital conversion, the very same configurations which enable us to recognize not only an o which is F, but also that o is F. Furthermore, the individual pieces of information isolated by each neural substructure in a definitional script can be identified with the frames of that scriptD. Each such substructure or frame would be responsible for indicating the presence of a single property of o, so that when enough of these substructures become “excited” or “stimulated,” object recognition (the recognition of o as F) is achieved (see §6.3 below).

This is a “bare bones” description of what a definitional script is. We shall later see that what gives scripts much of their cognitive power is their holistic potential—their ability to interact with and influence one another in a manner that makes them much more than mere sensors.

II. Because scriptsD are forged through the process of transforming information from analogue to digital form, they will—as all such conversions must—involve a loss of information. This loss of information is essential to the extent that the associated definitional script has the job of treating different things as the same, the job of classifying, or categorizing, or generalizing the input as a token of a general type.
That, I take it, is what concepts do. The suggestion here is that the definitional scripts we have of the things around us are the concepts we possess of those things. In order to form such scripts, we must ignore a great deal of what makes particular objects unique. My RADIODS includes many frames which collectively enable me to recognize radios when I encounter them. But what my RADIODS does not include is a frame indicating that Sony offers the “SACD stereo and multi-channel playback” as an optional feature for its DAV-C700 model. While this information may be part of my DAV-C700s (should I ever develop such a thing) it is certainly not necessary for my recognizing members of the class of radios as belonging to that class.

Humans categorize incoming information through the process of analogue-to-digital conversion, which is the essence of script formation. But although scripts greatly expand the human cognitive capacity, there is a trade-off: a loss in informational content means a loss in the discriminative power of the information-processing system, and the effect is proportional—i.e., the more data is lost, the more diluted a system’s ability to discriminate becomes, especially when the input is related to recognition. In fact, to prevent certain kinds of discrimination is precisely the point of classificatory organization, as we saw in the case of the speedometer and the four differently pitched tones.

This is a fairly straightforward idea: an information-processing system simply cannot categorize or classify things as types without somehow extracting the similarities and ignoring the differences between the objects involved. But “ignoring the differences” just is the selective disposal of information. What counts as “similarity” and “difference” is usually defined by the classification method itself. But it seems fair to say that any classificatory scheme will involve the removal of specific data, and this is certainly true of scriptal structures. We shall have occasion to return to this important result and its implications later. I simply wanted to mention it for now.

II. On the Precarious Existence of Water-Stains

The above sheds some light on the water-stains game of §4.1. Recall the rules: each player (myself included) was faced with a previously created watermark whose identity had to be determined to the satisfaction of the rest of the group; those who couldn’t make an identification were disqualified. A major difference between this childish diversion and the very serious business of object recognition is, of course, the reality of the “objects” involved, something about which we had no illusions. But that is just a matter of being in a particular “psychological mode,” to use Searle’s terminology. We still had to identify the shapes in the stains, regardless of whether or not we believed these shapes to represent “real” things.

The process by which we gave wet-patches their identities is essentially the same as the process by which people normally recognize objects in paintings, photographs, and even in real life: certain features or patterns in the water-stain (painting, picture, real-world object) arouse or activate or stimulate specific internal neural circuits that constitute part of some script. When that definitional script is triggered, it “guides our understanding” of the shape in front of us.

Even though water-stains were created without any attempt to manipulate their final shape, they may still be regarded as representations. The reason they may be so regarded is that we, the players, considered them as such. In fact, that was the whole point of the game, to try to determine the objects depicted by the stains. We were not always successful, of course, but our initial inclination was always to assume that the water-stains represented things. In short, we thought of them as pictures—not very good pictures, but pictures nonetheless. What I am suggesting is that
once water-stains were viewed in this way, the process by which we came to identify the objects represented by them was essentially the same as the process by which people in general come to identify objects in pictures and in the environment: *nothing is recognized unless a definitional script is primed or activated*. Depending on what that script characterizes, the object on the wall (painting, photograph, in the environment) is then "recognized" as a house, a turtle, a crow, or any one of countless other possibilities.\(^{12}\)

A failure to come up with a proper identification in the game was simply a failure to invoke any script at all. In such cases, the disappointed player clearly saw the water-stain, but did not see it as anything. Without taking the crucial step of associating the water-stain with a real-world object (based on perceived verisimilitude), water-stains were never elevated beyond their lowly existence as mere wet spots on a wall. They were not "transformed" into anything more substantial than dampness-caused discolouration.

On the other hand, there were times when a water-stain seemed to resemble more than one object. The threat of this happenstance was greatest when the watermark displayed a lot of "legs": anything which had four legs and a tail was a donkey; anything which had five legs had, in reality, four legs and a trunk, making it an elephant; anything with more than five legs—or four legs and a trunk, whichever way you want to look at it—was a spider. The trouble started when the players could not agree on whether the stain had four legs and a *tail* or four legs and a *trunk*. Disputes of this sort were rare, but when they surfaced longstanding friendships were often severely tested. These were trying times, indeed!

The ambiguity resulting from assigning watermarks multiple identities is easily explained as the priming of several distinct definitional scripts from the features of a single wet patch. The water-stain game is a good way to see this phenomenon at work, because the "images" we created were by no means precise, and the players thus had a lot of flexibility in associating them with common items. Photographs are different. An elephant in a clear photograph is unlikely to be mistaken for a spider because photographic images normally have enough *informational content* to activate a single, unique script. Generally, we may suppose that the less informational content emitted from a source, the more likely it is that interpretational ambiguity will accompany the analogue-to-digital conversion of this information.\(^{13}\)

---

\(^{12}\) The reason people know that they are looking at an object in a painting, or a photograph, or a water-stain rather than at something more tangible has to do with at least two factors: (1) there is the *mode of presentation* in which the object appears—i.e., the fact that it is a painting, or a photograph, or a water-stain we are looking at, and the fact that we recognize these representational modes for what they are; (2) there is *interactivity*, or what we can do with the objects in question. We can sit on a real chair, move it, kick it, take it apart and put it back together again, but we cannot do the same with a chair in a picture. With enough ingenuity, (1) can be circumvented: take a realistic, high-definition, real-size photograph of a chair, accentuate it with the proper lighting, shadow, and perspective, imbed it in a camouflaging environment, and you may fool a lot of people. But I don't see how (2) can ever be overcome.

\(^{13}\) A similar ambiguity once involved the identity of pandas. For a long time, scientists could not determine whether these animals belonged to the bear family or to the racoon family. The trouble was that many physical and behavioural characteristics of pandas were consistent with both species—which is to say that these characteristics did not discriminate between the **BEAR**\(_{es}\) and the **RACOON**\(_{es}\) of the specialists working on the problem—so an exact match proved elusive. It was not until identification was sought in more objective means that the issue was finally decided: genetic testing proved that pandas are bears.
iii. Complex Definitional Scripts

The concepts we have of inanimate objects, animals, and people—and the scripts that define these concepts—are, of course, different. We do not think of inanimate objects in the same way we think of living organisms, and within the class of living organisms, we do not think of animals in the same way we think of humans, strangers in the same way we think of friends, and friends in the same way we think of kin. The complexity of a definitional script depends on the kind of relationship the person whose definitional script it is enters into, or expects to enter into, with the object represented by that script: most of us just don't feel the same level of affection for rocks we do for siblings or parents, nor do we expect the same kinds of things from politicians we do from parrots (though perhaps we should). This is partly explained by the kind of stance we adopt toward such things (Searle 1983: 144).

I have tried to specify some of the main differences between the definitional scripts associated with disparate classes of objects in §4.4. I have argued, for instance, that scripts of living creatures tend to be more complicated than those of inanimate things, because (a) living objects grow and change, (b) living objects are more interactive than non-living objects and, therefore, (c) living objects tend to provoke emotional responses from us more frequently and with greater intensity. I have also discussed how scripts of persons can differ, as in the case of those for whom we have only the barest stereotypical notions (CHINESE-PERSONDS), famous historical personalities whom we only know through published reports or by word-of-mouth (CHARLES-DICKENSOS, WILKIE-COLLINSOS), figures of popular culture (PEE-WEE-HERMANOS, VANNA-WHITEOS), people who mean little to us beyond their functional role (the waitress in my favourite café, my dentist’s receptionist), acquaintances (coworkers, neighbours), friends (casual and intimate), family (immediate and extended), and sexual partners (husbands, wives, mistresses, sugar-daddies).

There is another difference between our scripts of animate and inanimate objects that should be noted. The uniqueness of the people we come to know (and perhaps some animals as well) tends to be their defining characteristic, unlike the common objects of daily life, which we normally mould into classes and types without caring much about what makes them special. If, as I have suggested, the dynamics of the relationships we develop with various entities is what largely shapes the sort of definitional scripts we form of these entities, then we may expect one’s script of an object belonging to one class of things to take on characteristics typical of a different class of things to the extent that the person starts to think of the object as an instance of that second class. An example of this kind of cognitive shift is often manifested in the relationship some children have with their toys: dolls, teddy bears, action figures, and the like. Not only do the physical characteristics of these objects become intimately known by their owners, but the objects also get well-defined “personalities,” with all the niceties such an endowment entails. On the other hand, we sometimes see the opposite shift take place. Sometimes we see humans (and animals) treated as artefacts, or at any rate downgraded to something less than they are. A clear example of this are the demeaning attitudes directed toward minorities in the propaganda campaign of Nazi Germany, eventually leading to many of the horrors of WWII.

But no matter how complex a definitional script is, no matter how intricate and untidy, it is still a necessary ingredient in the process of recognition. At this level at least, the idea remains simple and unaltered: we cannot recognize anything—whether it be living or non-living, an animal or a human, a stranger or a lover—without the presence of specific neural structures especially tuned.

14 These personalities are often suspiciously similar to their human counterparts.
to the features of the object in question. If these structures were lacking, we would be faced with the
same problem with which some of the water-stains players had to contend: we would see the object
in front of us, but not recognize it as anything in particular. So while there may not be a my-grandma
neuron, I maintain that there is a \textit{MY-GRANDMA}\textsubscript{DS}. Such a script would naturally be very complex,
including not only grandma’s physical features (facial characteristics, body shape, tone of voice) but
also her personality traits (kind-hearted, accepting, affectionate), likes and dislikes (hates rock music,
likes pomegranates), relationship with various family members (favourable to x’s company, can
barely stand y, indifferent to z), relationship with the person whose \textit{MY-GRANDMA}\textsubscript{DS} it is (as represented
by the memories that person has of their interactions), and other miscellaneous pieces of information
(what opera she went to see last year, her level of education, what year she married grandpa, etc.).

Of course, as long as people procreate there will always be grandmas whether we know it or
not. In that sense, grandmahood is an objective property whose possession does not depend on our
knowledge of it. Nevertheless, in order for Eva to recognize a certain woman as her grandma, she
must have a \textit{concept} of her grandma and apply that concept to the right woman, which, on my view,
is tantamount to Eva’s having a \textit{MY-GRANDMA}\textsubscript{DS} that is triggered or activated by the women who bears
the right relation to her. That is what it means for Eva to recognize the woman in question
as her grandma. Much of the explanatory edifice of cognitive psychology is based on the very basic human
ability to recognize things, and therefore much of this explanatory edifice presupposes—or at least is
consistent with—the theory of scripts.

\textbf{iv. Scripts, Scripts, and Scripts.}

We don’t just recognize artefacts and people. We also recognize many structured situations
as being of a certain generic kind or category. Our long discussion of episodic scripts in chapter IV
introduced us to the \textit{RESTAURANT\textsubscript{ES}}, the \textit{DOG-WALKING\textsubscript{ES}}, and the \textit{DOCTOR-VISITING\textsubscript{ES}}. Since all we have
to recognize in \textit{scripts\textsubscript{ES}} are the \textit{events} that constitute the episode, no linguistic indicators of any type
need to be present in order for recognition to take place. The following clarifies this point.

As Adam was flipping through the TV channels, he came across an old, silent, black-and-white movie that he had never seen before. Even though nothing about the movie seemed familiar (including the actors), he decided to watch it anyway. Eventually, he came across one segment consisting of the following three consecutive scenes (and let us assume that these scenes are independent of what came before and after them):

\begin{enumerate}
\item \textit{A man casually enters what looks like a commercial establishment.}
\item \textit{The same man is led to a table by a woman wearing an apron. After the man is seated, his lips start to move (but Adam could not make out what he was saying) just as the woman takes out a paper-pad and a pen from her pocket and starts to write. All around there are people sitting at tables eating and drinking.}
\item \textit{In the final scene the man is walking out of the establishment, counting his money, with a toothpick in his mouth.}
\end{enumerate}

\footnote{Such a script would be different for each person, of course, as grandmas differ from one person to another. I choose to call the script \textit{MY-GRANDMA}\textsubscript{DS}, but I trust it is clear that it may go by different headers—e.g., \textit{GRANDMA-GUIRGUIS}\textsubscript{ES} or \textit{GRANDMA-MARIAM}\textsubscript{ES} or \textit{GRANDMA-MARIAM-GUIRGUIS}\textsubscript{ES} or simply \textit{MARIAM-GUIRGUIS}\textsubscript{ES}.}
Did Adam understand what he just saw? Of course he did. Although no words were uttered, most people would have no trouble recognizing that the man in the segment has just had a meal at a restaurant—that he requested one or more dishes, received what he ordered, ate and paid for it, and left a tip for the waitress—even though these details were not depicted in the scenes. We assume these events because our RESTAURANTS guides our understanding of the situation, not because someone told us so. This suggests that whatever Intentionality a RESTAURANT has is not derived from (is independent of) the Intentionality of language.

Though Dretske acknowledges the importance of personal experience to learning in general, he does not specifically mention the learning of episodes—that is, the creation of scripts that comprise neural structures which have become sensitized to, and therefore acquired the job of indicating the presence of, specific features or properties of an ordered sequence of events. That such circuitry exists takes no great leap of imagination, and that it helps us recognize common episodes is evident each time we watch an old silent film or otherwise witness structured behaviour without the benefit of words.

What about knowledge of how to do things and knowledge of plans and motivations? We don't need language to recognize these either, and silent movies are also demonstrative of this fact. But even when language is present, it is sometimes more of a hindrance than an aid in identifying the nature of the personal and instrumental scripts others might harbour, especially when deception is practiced. At such times, we may expect language—used to express falsehoods—to draw us away from, not closer to, recognizing third-party scripts and scripts. A man preparing to commit a robbery, for example, would be a simpleton if he did not do everything in his power to conceal his intentions from others. When the time finally comes to carry out his deed, he will draw upon many well-practiced, largely unconscious instrumental scripts, like DRIVING-ACARs, PUTTING-ON-A-SKIMASKs, OPENING-ADOORs. But he will also draw upon more complicated personal scripts that require a lot of planning: e.g., how does one best achieve the D-CONT(money) delta-goal of the ROBBERYs? Does one use STEAL, THREATEN, or OVERPOWER? And how are these planboxes to be executed? A lot will be riding on the choices the man makes, including the ultimate success of his ROBBERYs and the kind of life he may expect to lead thereafter.

A Hierarchy of Physical Events Since personal and instrumental scripts embody large-scale schemata which breakdown into plans, goals, sub-goals, sub-sub-goals, eventually reaching down all the way to basic bodily movements (and beyond), the physical manifestation of these scripts must involve a multi-level hierarchy of physical activity, where only some of the strata are representational. One way to formulate this hierarchy is as follows:

\[ H: \text{ There is a hierarchy of levels of physical events, } L_0, L_1, L_2, \ldots, L_n, \ldots, \text{ of which at least one distinct level is associated with the subject matter of each special science.} \]

I echo Paul Humphreys' (1997: 5) observation that such a hierarchy would have to be an idealization of some kind: the interactions between the strata will be so entangled that any attempt to separate them would be difficult. However, our efforts will yield better results if we are aided by an appropriate division of labour. Separating the levels corresponding to the special sciences may be a job for researchers involved in these disciplines, and may depend on the basic working definitions of

---


17 It need not take any imagination at all, since brain-imaging studies have made that unnecessary; see §5.2.i.
these fields—e.g., what makes an event "(sub)atomic" "chemical," "biological," etc., and what, if anything, are the physical basis for such classifications. While we should not expect simple answers to these questions, we may suppose that as the sciences advance information relevant to the issues of autonomy and boundaries will become available. Meanwhile, we have the option of retaining an idealized hierarchy with no apparent negative consequences for the account proposed here.

Suppose, then, we accept \( H \), how shall we build our hierarchy? We must begin somewhere and so we start by positing a 0-level ("zero-level") as the lowest physical level of all, the level of fundamental particles and fields which govern everything else that happens. It is not important for our purposes to specify what kinds of particles and fields we are talking about, but what is important—what we are assuming—is that they do exist. A specific level \( L_i \) can now be defined as that level which includes all \( i \)-level events. 0-level events are defined in terms of the causal interactions of the fundamental particles and fields inhabiting the 0-level. A higher-level event may be defined in terms of its constituent lower-level events. Hence, event \( E_i \) is the set of all \( E_{i,1} \), events comprising \( E_{i,1} \), and an event \( E_{i,1} \) is the set of all \( E_{i,2} \) events comprising \( E_{i,1} \). So an event at the \( i \)-level will be identical to some proper subset of the entire corpus of events occurring at \( L_{i+1} \). These are identity relations which apply only to event-tokens. Relations among event-types \((E)\) must be characterized differently, using a basic supervenience principle: an \( i \)-level event-type supervenes on an \([i+1]\)-level event type if and only if there could be no difference in \( E_{i,1} \) without some difference in \( E_{i,2} \). The exact definition and nature of the supervenience invoked is a pragmatic matter, but a weak supervenience relation will probably do the job (Savellos and Yalcin 1995; Kim 1984).

Working "upward" from the 0-level, we can build intricate complexes of causal interactions corresponding to the familiar levels which form the subject-matter of the special sciences. The difference between any two levels will ultimately be a difference between the state of amalgamation and interactive forces among 0-level units, or clusters of such units, or clusters of clusters of such units, etc. Both the levels and their constituent events are separable systematically, and reflect a natural hierarchy of matter in various states of structural complexity. This architecture has the advantage of being as fine-grained as needed, for it supports many more distinct levels (and events) than might be required to distinguish between different types of explanation.

We can now explain a little better how the multiple nesting of goal and subgoal states may operate. Suppose you desire to acquire an object that belongs to someone else. We saw that D-CONT involves a basic set of ways (the persuade package) in which acquisition may be effected: ASK, INVOKE THEME, INFORM REASON, BARGAIN OBJECT, BARGAIN FAVOUR, and THREATEN. We may expect, therefore, that the general ACQUISITION will have the members of the persuade package among its frames. Whether or not any of these members is ultimately used to discharge D-CONT is something that depends on the context of the situation and the nature of the actor. In any case, there are liable to be more frames in ACQUISITION than those defined by the persuade package—such as, for instance, STEAL and OVERPOWER. To keep things simple, we shall assume that the set of frames of ACQUISITION is exhausted by the persuade package plus STEAL and OVERPOWER, designated "PP+SO" (Figure 6.1).

Just as (some of) the levels of Figure 6.1 may be thought of a psychological insofar as they are implicated as reasons in the explanation of actions, so too might these levels be thought of as physical insofar as they embody neural circuits with specific representational chores. Hence, ACQUISITION would be physically implemented in the form of a large neural structure which plays both an indicatory role in recognition and a guiding role in behaviour. Since the acquisition script is a broad concept that includes its subgoal states as cognitive units, we have good reason to suppose that ACQUISITION in physically realized at a higher level than those of PP+SO, and that the physical realization of ACQUISITION includes the different circuits which form the physical bases of PP+SO. In
turn, all of the planboxes of PP+SO will be realized in the brain at higher physical levels than the levels of their subordinate subgoals.

\[ L_{i,j > i} \quad \text{ACQUISITION}_s (D-\text{CONT}) \]

\[ L_{j,i > j > k} \quad \text{persuade package} \quad \text{STEAL} \quad \text{OVERPOWER} \]

\[ L_{k,j > k > m} \quad \text{ASK} \quad \text{INVOK Themes} \quad \text{BARGAIN FAVOUR} \quad \text{BARGAIN OBJECT} \quad \text{THREATEN} \]

\[ L_{m,k > m > n} \quad \text{USE KNIFE} \quad \text{USE GUN} \quad \text{USE BODILY FORCE} \]

\[ L_{n,m > n > 0} \quad \text{Muscle Movements} \]

\[ L_{0,0 - \text{level}} \quad \text{Neural Activity} \]

\[ L_{0,0 - \text{level}} \quad \text{Fundamental Particles and Fields} \]

Figure 6.1: A representation of the ACQUISITION hierarchy. Starting at the 0-level of fundamental particles and fields, we can work up to conscious plans and goals. Some of these levels, particularly those invoked as reasons for actions, can be used in psychological explanations—e.g., the wallet was lost because the thief scared the owner, and the thief scared the owner by putting a knife to his throat.

To illustrate, ACQUISITION specifies that OVERPOWER is one possible way of achieving D-CONT, so ACQUISITION constitutes a higher-level neural structure than that of OVERPOWER. Moreover, the neural circuits which represent OVERPOWER will be a part of the neural circuits which represent ACQUISITION. Similarly, OVERPOWER specifies different methods of using force—e.g., with a knife, a gun, or raw physical strength. These subgoals will therefore be implemented in neural mechanisms operating at a lower level than (and constituting a part of) the more complex neural circuits representing OVERPOWER. We continue this imbeddedness until we get to levels of physical activity that have no explicit representations at all, eventually reaching the "rock-bottom" strata of fundamental particles and fields.

The multiple nesting of (sub)goals distinctive of scripts and scripts make it difficult and somewhat arbitrary to distinguish between these kinds of scripts and the frames that constitute them. Are the planboxes of PP+SO the frames of ACQUISITION or are they instrumental scripts in their own right? I doubt any answer will be entirely satisfactory, partly because these planboxes can take on both roles. When ACQUISITION is the focus, members of PP+SO are best viewed as frames, but when the focus is on one of these members—e.g., STEAL—it should be considered a separate script with its own set of frames. Such a conception suggests something we have already been assuming, that
scripts may come in grades, where we might have a script within a script within (possibly) more scripts. There is no in-principle objection to looking at things in that way, just as long as we keep track of all the elements involved and realize that the regress will eventually end at the level of neuronal firing.

Note, finally, that scripts and scripts are similar to all other scripts in that they play a double role. They are involved in behaviour, where they guide us in achieving our wants, and in cognition, where they help us to recognize and understand the wants of others. Moreover, our hierarchical explanation of how these scripts may be physically implemented does in no way undermine or challenge the use of these scripts in psychological explanations. Nor are neural structures barred from manifesting psychological effects merely in virtue of the fact that their operations can be described at the physical level. This assumption is in keeping with a realist-cum-materialist attitude toward mental states. Dretske has already told us why there is no tension between physical and psychological explanations: accounts in terms of reasons tell us why an action came about; neurophysiological reports tell us how it was accomplished.

6.3 Priming by Autoassociation

It is one thing to say that recognition-comprehension is what we get when a specific script is applied to some object or situation, but a different thing to say how this “applying” works. What is it that activates the when we see a movie segment like the one Adam saw, or the when we come across chairs in all kinds of places? Which frames of an object are necessary for priming o’s corresponding script? Which are only sufficient and which are both? Can we even make a determination?

1. The Hebb-Marr Model of the Hippocampus

One of the earliest and most influential theories of hippocampal processing is relevant to these questions. The theory was proposed by David Marr (1971) who, starting with what was known then about brain anatomy, sought to infer an emergent information-processing capability of the hippocampus. His ideas gave rise to a broad class of accounts, often termed the Hebb-Marr models because they incorporate Hebb’s (1949) views on how associations are acquired between groups of cells in the brain. Since Marr’s original publication, new empirical data have shown that some aspects of his analysis are incomplete or incorrect (Willshaw and Buckingham 1990). But these aspects are irrelevant to the discussion that follows. Marr’s central notion of autoassociation has withstood experimental testing and remains the basis for current models and theories of the hippocampus.

Marr’s basic idea was to distinguish separable roles in memory for the archicortex, including the hippocampus, and for the neocortex. He assumed that the chief role of the neocortex was to

---

18 The cortex of the mammalian cerebrum admits a crude division into two regions: the archicortex, which is relatively simple and primitive, and the neocortex, which is evolutionarily more recent and very elaborate, especially in humans. The distinction between archicortex and neocortex is thought to reflect a difference in their functions. The archicortex is essentially a memorizing cortex, whereas the neocortex is regarded as a classifying cortex. Marr’s analysis of the hippocampal formation was based on his studies of mice, rats, guinea pigs, and rabbits. But as he points out, “there is ... a remarkable uniformity in the structure of the hippocampus in mammals ..., so that the divisions made in the mouse are easily recognizable in man. The only important histological difference is in the size of the elements involved: man’s hippocampus is larger in every way than that of the mouse” (1971: 55).
store large complex event memories—broadly equivalent to what we have been referring to as episodic memories—composed of several integrated associations. For example, the event memory of a meal might include associations about the food eaten, the meal’s location and time, and the company sharing it. In Marr’s model, an event memory is defined as a pattern $E$ of activities over a large number of neocortical cells, evoked by a particular set of sensory inputs. Such a pattern is stored by associating its elements so that activation of some of the cells representing elements in $E$ can activate other elements in turn. Later, if a subset of $E$ is presented to the neocortex, the neocortex should be able to retrieve the full pattern. This ability, called autoassociation, is illustrated in Figure 6.2 (adapted from Gluck and Myers 1997).

![Figure 6.2: Storage of an event memory as a pattern of cell activations in the neocortex. (A) Initially, the event memory simply evokes a pattern of activations (darkened circles) across a group of unrelated cells. (B) As the pattern is stored, various elements of the pattern are associated by weighted connections (lines). (C) Later, if a partial version of the original pattern is presented (darkened circles), activation spreads along the associations to activate the complete pattern (D).](image)

One difficulty in implementing this function in the neocortex is that a large number of very precise connections are required to associate the elements in $E$ with each other. Further, the associations required to store $E$ might well disrupt pre-existing networks created to store other patterns with common elements. Worse, if another stored pattern $F$ shares common elements with $E$, then $F$ may interfere with attempted retrieval of $E$: if a subset of $E$ is presented, activation will spread to these common elements, which will then begin to retrieve $F$ as well as $E$. At the extreme, if many overlapping patterns are stored, an attempt to retrieve any one of these will result in an array of
activation that shares elements with all stored patterns but is identical to none. This situation is called catastrophic interference.

Because of this potential for interference in recall, Marr suggested that it would be useful to have a separate processor—such as the hippocampus—that could rapidly store event memories and then allow gradual transfer of this pattern to the neocortex. The neocortex would then reorganize and classify this information, situating it within existing knowledge and memories. More specifically, Marr proposed that the hippocampus is able to rapidly store new patterns, holding them in a temporary memory store, but is not able to integrate them within the larger body of existing information.

He imagined the hippocampus as functionally consisting of two layers or groups of cells. Inputs cause activity on the first A layer, which projects onto the second B layer. The B cells then project back to the A cells. All synapses between cells are modifiable, but they are simplified to allow only binary “on” or “off” values. This network is the same as that shown in Figure 6.3. A stored pattern can be retrieved as follows: when part of the pattern is presented to the A cells, the evoked activity on the B cells feeds back to complete the original firing pattern on the A layer. This autoassociation process forms the basis for many current and more detailed theories of hippocampal physiology and function.

Autoassociation is essentially how scriptal priming works. But, of course, what we have described here is a very simplified version of the phenomenon. Most of the scripts an adult possesses are much more elaborate than the configuration of Figure 6.3, and will probably involve connections that span over a number of distinct functional regions in the brain. But the basic idea of autoassociation is the same in all cases and must be involved in many cognitive processes—like comprehension, memory retrieval, and the kind of plasticity that allows us to apply concepts in a variety of different and novel ways.

Object recognition is one case in point. As mentioned, to recognize o as F is simply to apply $F_{DS}$ to o, and the application of $F_{DS}$ to o is a matter of o’s priming $F_{DS}$. Priming occurs when some of o’s physical characteristics stimulate those neural circuits (constituting at least a part of $F_{DS}$) whose function is to detect the presence of the characteristics in question. The entire script becomes active through autoassociation: the stimulated neural structures are a part of the A layer, which receives direct input from the external environment—in this case, o. The A structures connect with a second B layer of cells in a manner that permits bi-directional influence. When only a proper subset of the entire number of frames comprising $F_{DS}$ is presented to the A layer, the B layer feeds back and activates the remaining neural structures in A to complete the invocation of the stored script.

This model of scriptal priming provides fairly simple responses to the questions we asked at the beginning of this section: which frames of $F_{DS}$ are necessary, or just sufficient, or both necessary and sufficient for its priming? Answer: despite the fact that people are generally very skilled and consistent at recognizing things, there is no single frame that is either sufficient or necessary for

---

19 Marr (1971) suggested that the fraction of a previously stored pattern needed to successfully retrieve the full pattern from store is $\frac{1}{3}$. At the time Marr proposed his theory, computer technology was not available to simulate neural nets at all. Even today, technology does not allow the simulation of systems anywhere near the size analyzed by Marr. Nevertheless, Willshaw and Buckingham (1990) undertook a computer analysis with a more modest system. They found that partial cues of 8% of the size of the full stored pattern resulted in 66% retrieval, and partial cues of 25% resulted in perfect recall; 99% retrieval was obtained from 16% cue size. Marr, it seems, was a little off, but not by much.
priming a script—a fact which actually helps us to explain why humans are generally proficient object recognizers. How do we know that \( o \) is, say, a chair when,

![Diagram](image)

**Figure 6.3:** A schematic of Marr’s model of the hippocampus. Cells are either \( A \), which receive direct activation from external input (heavy arrows) or \( B \), which are driven only by \( A \) cells and affect them in turn. Learning consists of strengthening the connections between \( B \) cells and the \( A \) cells that activate them. Later, if a partial version of a stored pattern is presented to the \( A \) cells, the \( B \) cells feed back and activate the remaining \( A \) cells required to complete the stored pattern.

- we see and recognize chairs under numerous lighting conditions: natural and artificial, bright and dark, and in multi-hued illumination;
- there is no standard material out of which chairs are composed: they come with steel, wooden, or plastic frames; they are upholstered using a variety of natural and artificial fabrics, or sometimes not upholstered at all;
- chairs have no standard shape: some have armrests, some don’t; some have four legs, some have more or less; some have long, wide backs and some have short, thin ones; some are high, some are low; some are big, others are small;
- there are “office” chairs, and “lawn” chairs, and “dining” chairs, and even “wheel” chairs.

The fact that \( o \) can manifest any combination of these properties but is still recognized as a chair would be very difficult to explain if there were singularly necessary and/or sufficient properties for priming \( \text{CHAIR}_{os} \).

20 This does not mean that \( \text{CHAIR}_{os} \) has no default frames, since default frames are not necessary in any sense. They are the features or properties we assume when there is no explicit information given one way or the other—e.g., as in the case of occlusions. One’s default frames of \( \text{CHAIR}_{os} \) will probably be the stereotypical features (s)he associates with chairhood, and these will vary across individuals depending on personal experience.
Suppose, instead, that the requirement for activating a script is that the stimulation of the $A$ layer of neural structures only exceed a particular threshold. When this threshold is reached, the $B$ layer then takes over and activates the rest of the elements in $A$. In this case, none of the frames (i.e., neural circuits) of $A$ is likely, by itself, to be either adequate for this activation or necessary to produce it. What counts only is that some indeterminate combination of these frames does the detection work it was designed to do. Which combination exactly will vary depending on the perceptual situation and the nature of the object perceived. But as long as the stimulation of the $A$ layer exceeds the relevant threshold, the entire script becomes active and recognition takes place.

The biggest advantage of this hypothesis is the incredible versatility it gives an organism. No two instances of a single type need be exactly the same in order to be recognized as instances of that type. And this gives us a way to answer—not evade, but answer—the questions with which we started this section. As far as recognition is concerned, there are no essential qualities of objects, and sufficiency is defined by the level of stimulation required to activate a script. Thus we become acquainted with many different instances of chairs through experience, and this allows us to create a reasonably rich $\text{CHAIR}_{os}$ that includes many of the peculiar features these instances have displayed (colours, materials, sizes, shapes). But depending on the physical properties of the chair we see at a given time, only a subset of these frames will become stimulated. Whether or not the $\text{CHAIR}_{os}$ is subsequently elicited depends on whether or not the stimulation level exceeds a critical value. In the same way, we can recognize chairs even when they display never-before-seen characteristics, just as long as their more "typical" features stimulate the sub-circuits of $\text{CHAIR}_{os}$ beyond its activation point.

The same idea works for all other objects. A cow may formally be a large quadruped, but we would have little trouble dealing with a three-legged cow amputee, as long as it is a reasonably good cow in most other respects. If a person is shown something that fits none of her stored scripts, her identificatory system—whose job is to find a match of some sort—will at least point to an object the sample is "close to," along with some indication of the major deviations from that object. An identificatory system organized along these lines might easily come up with “Like a man, but with pointy ears and green skin,” or “A woman from the waist up, and a sardine from the waist down.”

ii. Priming by Description

I have claimed in §4.4.i that scripts may be informed indirectly as well as directly: not only are our scripts influenced by what we see, hear, taste, touch, and smell, but also by linguistic information sources, such as stories, books, magazines, and the like. In fact, linguistic sources are probably the most important way in which learning is achieved. I have also said that scripts are just as primeable by words as they are by non-verbal cues.

In order to see the connection between these two assertions, I want to draw attention to a peculiar feature of natural language, one to which Dretske has alluded in his discussion of the digitalization of analogue information:

If I simply tell you, "The cup has coffee in it," this (acoustic) signal carries the information that the cup has coffee in it in digital form ... If, on the other hand, I photograph the scene and show you the picture, the information that

21 Note the emphasis on "recognition." I am not claiming that there are no essential properties of things; the property of being an animal seems to me to be essential to the property of being a tiger. But I do claim that no single property—not even that of being an animal—is essential for the application of my tiger script to a given object. So I may apply my $\text{TIGER}_{os}$ to a realistic tiger-like robot, thus mistakenly but sincerely believing that the object I see is the actual article.
the cup has coffee in it is conveyed in analogue form. The picture tells you that there is some coffee in the cup by
telling you, roughly, how much coffee is in the cup, the shape, size, and color of the cup, and so on. (1981: 137)

Dretske uses vision to make his argument, comparing a picture of a cup of coffee with a linguistic
description of the same. The picture displays a lot of "unnecessary" information—unnecessary, that
is, if the objective is just to show that the cup has coffee in it—thus making it an analogue
informational source. The natural language statement, on the other hand, "cuts through" all the
fortuitous particulars and simply relates the relevant data digitally.

Notice, however, that the linguistic statement could have been a lot longer, long enough to
include many of the details the picture portrays. In other words, instead of the simple and precise
"The cup has coffee in it," we could have chosen to write a multi-page study describing in as much
detail as possible the image in the picture. Both the short and long reports would carry the
information that the cup has coffee in it, but the latter comes much closer to being in analogue form
than the former.

Notice also that Dretske could have chosen one of the other senses to make his point. I can,
though perhaps with a bit more difficulty, describe to you in words what a particular texture feels
like, or a particular odour smells like, or a particular flavour tastes like. Of course, it is no accident
that Dretske opted for vision, since that mode of sensation is particularly poignant, and seems more
amenable than the rest to linguistic expression.  

The idea to which I want to call attention is this: Dretske takes for granted that both the
picture and the verbal message carry the information that the cup has coffee in it. Even though the
picture carries the information in analogue form and the verbal (acoustic) signal carries the
information in digital form, the information that the cup has coffee is still conveyed in both ways. I
can also use verbal expressions (signals) to convey information about an odour I smell in the kitchen,
or the distinctive sound of a tuba, or the flavour of a specific herb, or the texture of some surface. In
the same way that the picture of the cup of coffee and the sentence "The cup has coffee in it," convey
the same piece of information (albeit in different forms), language can also express the same sort of
information the other sensory faculties do. All of this suggests a possible idea about the original
function of natural language as a kind of complement to the senses. Call it the language-as-sensory-
complement hypothesis, or LASCH for short.

**LASCH**

**Preamble:** Many animal species (including humans) possess sensory tools—vision, smell, hearing, taste, and touch—to facilitate survival and to cope with their changing environments. But people have the sole distinction of having also developed, through the process of natural selection, a genetic capacity to communicate in natural language.

---

22 Perhaps this is why literary authors are often successful at describing powerful scenes in prose and poetry. We are moved by such descriptions to the extent that they are vivid and ornate. But sometimes immense skill is reflected in an author's ability to convey truly compelling images in so few words. A good example of this is William Carlos Williams's (1938) poem, "Young Woman at a Window."

23 Not all animals enjoy all of these faculties. The cobra, for instance, has no auditory sense, and the earthworm is blind. On the other hand, some animals have all five senses plus more—e.g., the echolocation capability of bats. We ignore such deviations here.

24 The claim is that there is a genetic basis for the human capacity to learn natural language, but not, of course, **English** or **Swahili** or **Urdu**.
Chapter VI  Thoughts on Physical Implementation

Hypothesis: This evolutionarily recent ability gave members of the human species another way to process and express essentially the same information that is typically processed by the five senses. What our ancestors saw, heard, smelled, tasted, and touched, they were also able to articulate in words—no doubt very crudely in the beginning, but eventually in as much or little detail as they pleased. On this hypothesis, then, we can think of the original function of language as an informational complement to the senses. Whatever may be the evolutionary facts about how this ability developed, it was highly advantageous for our early ancestors. It also explains many current human cognitive abilities.

LASCH is highly speculative and I shall make no serious attempt to defend it; much of its plausibility derives from its explanatory potential, to which I will turn shortly. I suggest the hypothesis for two reasons: (1) as a way of putting some perspective on the linguistic priming of scripts; and (2) as a means of explaining how we acquired the ability to form scriptal representations of abstract or fictional objects. Many of the things I shall say about language in this regard may seem commonplace, but that should only serve to make LASCH all the more plausible. Before we get to these details, a couple of points bear clarification.

First, to say that language can convey the same data as our sensory systems is not to say that this data is conveyed in the same way. Adam may see Eva coming or he may smell her coming. The information that Eva is coming is carried in both instances, but it is being carried by different kinds of signals. There is also the matter of the form the signal can take. Dretske tells us that a picture of $o$ and a (concise) linguistic statement describing $o$ can both carry the same information about some aspect of $o$'s state, but the first does it in analogue form while the latter does it in digital form. If we let the word “mode” denote the different ways in which the senses impart information, and the word “form” denote the analogue-digital capacity of a signal, then the same information can be conveyed in different modes as well as in different forms.

Second, to say that we might evolutionarily conceive of natural language as an informational complement to the senses is not to say that this is the only or even primary function of language as it is currently employed. How and why we use words nowadays may bear no resemblance to, nor retain any of the original “purpose” for which the capacity to use words developed long ago.

Now, what kind of cognitive power does LASCH bestow upon those organisms happy enough to use words? Here is a partial answer:

1. LASCH gave our ancestors the shortest of shortcuts. Not only did it enable them to convey sensory information in a different way, but it also allowed them to do so by organizing certain facts (about prey, predators, tools, environmental conditions, or whatever else was useful to them) into “bundles” attached to labels or headers (“antelope,” “lion,” “blade,” “rain,” etc.). These “bundles” constitute what we might consider today fairly simple concepts or scripts.

2. Once we got labels, an entire universe of possibilities started to open. We became more and more adept at communicating large volumes of information quickly and efficiently. We started to extract the common properties of things and to give these headers as well. We started to manipulate concepts beyond their “common” roles, flirting with new conceptual

---

25 Assuming that Adam has an uncanny sense of smell that is at least as informationally precise as his vision.
possibilities and combinations. As these experiments continued, language started to become more than just a conveyer of sensory information; it started to become the vehicle of abstract thought, and new kinds of scripts—related to events, plans, motivations, goals—started to take shape.

3. The headers given to scriptally-organized information are the best and most direct way to prime or invoke the information in question. But since language can carry the same data as the senses (only in a different way), the right sort of linguistic description can produce the same priming effect on the relevant script as an actual corporeal object, otherwise the whole point of having a linguistic informational complement to the senses would be lost. So whether the information is imparted verbally, visually, or in some other way, as long as it is the same information, it will prime the associated script.

4. LASCH gives us an account of how indirect knowledge is acquired. If language developed to complement the senses, then we can learn about things—we can form all kinds of scripts—without necessarily having direct sensual experience of the things learnt. We can learn about remote places and countries, about distant people and people who are no longer living. We can consider counterfactuals described in words as if they were real states of the world and contemplate their consequences. We can learn what to do and how to behave in hypothetical situations, and reflect on moral problems. The reason we can do all of these things is the idea that natural language can convey the same information as the five senses without being confined, as the senses usually are, to what is true or spatiotemporally local.

5. LASCH explains imagination, our ability to create fictional people, places, and events. Because we do not have to actually experience objects through the senses in order to build scripts of them, we can easily make a Santa Claus and a Tooth Fairy. We make Leprechauns, Werewolves, Vampires, Pixies, and Hobbits. All of these “creatures” get definitional scripts—complete with physical, social, and behavioural characteristics—just as other living things do. The only difference is the mode in which the relevant data is transmitted plus the fact that the ontological status of these creations is known to most people. There is thus a close kinship between imagination and indirect learning, for similar enabling processes underlie both phenomena.

6. Scripts of actual and fictional objects can now interact unencumbered, for that is partly what it means to have an imagination. So we can imagine a fire-spitting dragon having a conversation with Richard Nixon about a topic of mutual interest! And we can sing about situations where mommies are “seen” kissing Santa Claus. We can also imagine two historical figures interacting in some way based on what we scriptally know of their personalities and convictions. Though the individuals themselves would not be invented, their encounter would. What we need to conceive of such things are well-defined scripts (whether of real or manufactured objects) and the ability to manipulate these scripts in various ways. In most cases, we indulge in such reveries without much confusion, since our scripts tell us what is real and what is not. On the other hand, a child may take Santa to be just as genuine as the kiss he saw him give his mother. The child, is this case, would be in error, a topic we shall discuss in more detail in the next chapter.

7. LASCH counts among the most important differences between humans and other animals. Lacking a natural language, non-human species are incapable of being informed—or, at any rate, are severely hindered in that capacity—by any source other than their senses. They
cannot, therefore, ponder hypothetical or contrary-to-fact possibilities. And this strictly limits the range and application of whatever concepts they may possess.

These ideas, I admit, are not very deeply developed, but I trust the point is made. We have in language something that can take over any, or indeed all, of the informational functions of the senses. Language cannot convey the purely phenomenal aspects of experience, of course, but it does have the capacity to convey the same information—just as one can either see, feel, or be told that a cup has coffee in it. This kind of informational overlap allows us to “break free” from the senses and reflect on things that we cannot directly perceive in a way that other animal species can never do. This ought to satisfy those who insist that natural language is essential for the kind of cognitive power that is a part of being human. For LASCH has it that abstraction and imagination—two crucial ingredients in higher cognition—were facilitated by a functioning ability to use words and would probably not have been achieved without it.

6.4 Original Intentionality, Holism, and the Frame Problem

A few more things need to be said of the account sketched in this chapter before we get to the philosophical consequences of script theory. First, scripts are wholly intrinsic natural indicators created and used by agents. They are, in other words, Type III representational systems with real contents, and therefore, have reference (the objects in the environment they pick out) and sense (the way in which these objects are represented). Moreover, as Type III systems, Dretske’s argument regarding original Intentionality applies to them. The Intentionality of scripts is thus “self-made” and not bestowed in the manner that, e.g., symbols acquire their Intentional properties.

The next issue is a bit more complicated. Dretske’s information-processing account of cognition has been criticized for being a little too simplistic. It appears to require only very basic associative learning of the type actually done by PDP networks and sea-slugs. To get to real beliefs, we may need to add provisions concerning holistic systems of representation. This is something that Dretske had anticipated. What enables our beliefs to float freely, says he, is just “the interdependent, holistic character of meaning in mature, tightly integrated representational systems” (1988a: 138).

Dretske uses the notion of explicitness to draw our attention to the holistic character of the human conceptual network. An explicit belief differs from an implicit belief with the same content in that “the content of ... [implicit] beliefs is not available, as it is with explicit beliefs, for other jobs” (1988a: 118). Thus, drawing on a (1986) discussion by Crispin Wright, Dretske suggests that a rat’s belief that $x$ is poisonous may have only a single form of expression in the rat’s behaviour: the rat avoids $x$. In contrast, a human belief that $x$ is poisonous may manifest itself in an indefinite variety of ways. Wright elaborates:

I may, like the rat, avoid the substance. But I may also take steps to ensure my family avoids it, or take steps to ensure they don’t, ... I may take a large quantity if I wish to commit suicide, and a smaller one if I wish to incapacitate myself so as to avoid an obligation. My belief that the substance is poisonous is thus, as Evans puts it, at the service of indefinitely many potential projects corresponding to indefinitely many transformations in my other beliefs and desires. With the rat, in contrast, concepts like the desire for suicide, or malign intent, can get...
no grip ... [It]s 'belief' that the substance is poisonous has consequently no other expression than in shunning it. (1986:33-4)

This is ultimately what enables us to link our beliefs “into cognitive and conative molecules that exhibit some of the organizational and explanatory complexity of our reasons for acting” (Dretske 1988a: 122).

Dretske maintains that his information-processing theory is compatible with cognitive holism. His basic account “identifies what we believe with what it is the function of certain elements to indicate” (1988a: 150). So whether or not a neural structure represents anything will depend on whether or not that structure has acquired the job of indicating or explicitly storing knowledge related to some object, procedure, goal, or event. Now imagine a more complex system in which an internal state, which once indicated that $F$, can come (through further experience) also to indicate that $G$. Thus an internal state $F$ may indicate a particular call made by Meerkats, and another state $H$ may indicate the presence of a predator. If an animal comes to recognize the call as a warning cry, the state $F$ has come to indicate the presence of predators also. In this way elements acquire additional indicator functions and so change their meanings (given that meaning is fixed by indicator functions, which is Dretske’s basic claim).

This explains why integration into a web of concepts can alter semantic properties—i.e., it explains why holism matters. A child and a pigeon may each be equally good at discriminating trucks. But children—partly because they have a handy linguistic sign (label, header) to keep referring to—are able to associate truckhood with many other things: that trucks are solid objects that move, require fuel, run on highways, and so on. And that, Dretske says, is in part why the pigeon lacks our concept of truck: it lacks the body of collateral information to which we are privy.

As beliefs become integrated into more tightly structured cognitive systems, their indicator functions become more interdependent. Not having to do as much, they are free to be more specialized. As a result of this increasing specialization, they begin to exhibit a finer-grained Intentionality. This, in a nutshell, is why a rat’s belief that a light is on might differ from my belief that the light is on while remaining, in an important sense, a belief that the light is on. (Dretske 1988a: 150)

Cognitive holism is not something that I merely tolerate; it is something upon which I insist. At the end of chapter III, I claimed that as people seem to recognize objects and judge situations mostly by sensory cues, an account of recognition and comprehension—and eventually of narrow psychological content—must involve some kind of descriptive story. By “descriptive story” I meant that the physical, functional, and/or behavioural characteristics of an object $o$ will often have a role to play in activating a related script of $o$, and will thus figure in our recognizing or comprehending what $o$ is. Much of the above has been my attempt to sketch how such priming takes place.

I have also claimed that our descriptive story will require some measure of holism which, far from being a weakness, is actually in keeping with the way in which the human cognitive system seems to be organized. But I am inclined to think that Dretske’s explanation of holism does not go far enough, partly because he considers only one way in which holism can come about—namely, when an internal state changes its meaning by acquiring additional indicatory duties. The problem is that this process does not address the kind of holism derived from the nesting of concepts.

Consider again our old friend, the RESTAURANTES. It is an episodic script that tells us what happens in a restaurant situation: the sequence of events, the people involved, the props used, and so forth. But if this is the case, then it appears that our generic RESTAURANTEs must now include, as a part of the scriptE itself, additional kinds of scripts—e.g., definitional scripts like WAITEREs, TABLEEs, and
thoughts on physical implementation

SPOONos. Things only get worse from here, since each of these nested scripts encompasses or connects to its own unique set of concepts. For example, there seems to be a link between WAITERos and HUMANos in that the former is an instance of the latter. There is also a connection between HUMANos and MALEos or FEMALEos, and between these last two and REPRODUCTIONos, SEXos, PLEASUREos, and on and on. What we end up with is a messy network of ideas, where each is either linked to other ideas or forms a part of a bigger schema.

And then there is the frame problem, which appears in AI attempts to represent a shifting environment. Eric Lormand (1998) tells us that the original frame problem is somewhat different from the philosophical concerns that have been subsequently raised:

From its humble origins labelling a technical annoyance for a particular AI formalism, the term “frame problem” has grown to cover issues confronting broader research programs in AI. In philosophy, the term has come to encompass allegedly fundamental, but merely superficially related, objections to computational models of mind in AI and beyond.

The classic frame problem appears within the situation calculus, a language of predicate logic used to represent and reason about action and change. We can think of the situation calculus as a kind of “knowledge representation” code; it was originally developed by McCarthy and Hayes in 1969, and is commonly used in artificial intelligence for purposes such as predicting the effects of actions on a system’s state, planning actions to achieve given goals, diagnosing what events might explain some observations, and analyzing the operation of programs that perform such tasks.

Typically, in such systems there are “deductive axioms” about changes that are conditional on prior occurrences—that pressing a switch changes the illumination of a lamp, that selling a car changes its ownership, etc. Unfortunately, since inferences are made solely by deduction, axioms are also required for non-changes—that pressing the switch does not change the owner of the lamp, that selling the car does not change the illumination level of its headlights, etc. Without such “frame axioms,” a system is unable to deduce that any states persist. The resulting problem has to do with the potentially astronomical number of frame axioms needed to relate each representable occurrence to each representable non-change.

Some philosophers of mind believe that the original frame problem foreshadows deeper difficulties for the AI ambition of building interactive computer programs, or at least for cognitive science more broadly construed. For instance, Daniel Dennett (1987c) wonders how to ignore information obviously irrelevant to one’s goals, as one ignores many obvious non-changes. John Haugeland (1987) wonders how to keep track of salient side effects without constantly checking for them; this includes the so-called “ramification” and “qualification” problems of AI. And Jerry Fodor (1987b) wonders how to avoid the use of “kooky” concepts that represent intuitive non-changes as legitimate changes—e.g., “fridgeon” which applies to physical particles if and only if Fodor’s fridge is on, so that Fodor can change the entire universe simply by unplugging his fridge. Some AI

27 We never get scripts within scripts within scripts ad infinitum, and it has been the burden of this entire chapter to show why. In accordance with the account of concepts given by Dretske, scripts are formed as a result of the analogue-to-digital transformation of data, and the content of a script arises as the neural structures constituting the script acquire indicatory-identificatory duties during the process of this transformation. So there is no regress problem, either for me or for Dretske.

28 The full text of Lormand’s entry is found at: http://www-personal.umich.edu/~lormand/phil/cogsci/frame.htm.

29 One approach avoids the use of huge numbers of axioms about non-changes by using implausibly bold axioms. For example, one tactic is to assume that all the possible causes of a certain kind of effect are familiar, or that all the actual events or actions operating on a given situation are known (Lormand 1998).
researchers, including Drew McDermott (1987) and Pat Hayes (1987), protest that these “philosophical” issues are unconnected to the original frame problem.\(^\text{30}\)

Nevertheless, as Lormand (1998) points out, the philosophical challenges must be met if the complexity of human cognition has any hope of being mirrored in computational terms. The core of the philosophical difficulty has been illustrated nicely by Dennett (1987c) with a story about a robot named R\. The only task for the robot was to fend for itself. One day the designers arranged for it to learn that its spare battery was locked in a room with a time bomb waiting to go off soon. R located the room and formulated a plan to rescue the battery. There was a trolley in the room and the battery was on the trolley. R hypothesised that a certain action—which it called PULLOUT—would result in the extraction of the battery. It acted quickly and did indeed remove the battery before the bomb exploded; sadly, the bomb was also on the trolley. R had failed to realize that pulling out the trolley would also bring the bomb with it. The designers realized that the next robot must be made to recognize, not only the intended implications of its actions, but also the implications of their side-effects. They called their next robot, R\(^D\), the robot-deducer. When R\(^D\) was faced with the same predicament as before, it also decided that it should pull out the trolley to retrieve the battery. It had just finished deducing that pulling the trolley out of the room would not change the colour of the room’s walls when the bomb exploded. The designers realized that they must teach the next robot the difference between relevant implications and irrelevant implications. Thus, R\(^D\), robot-relevant-deducer, was faced with the same dilemma. The designers were surprised to see R\(^D\) sitting outside the room pondering its actions when the bomb exploded. When asked to explain itself, R\(^D\) responded that it was busily ignoring thousands of implications that it had deduced to be irrelevant.

All of these robots suffer from the frame problem. Dennett (1987c: 47) remarks that the reason AI forces the trouble to the surface is because, in attempting to build an intelligent machine, researchers must “start at zero.” A computer program has to be told everything; things that we (as intelligent beings) take for granted. We know literally trillions of things—e.g., that a brick is heavier than a feather, that knives are normally sharp, that opening a fridge door doesn’t cause a nuclear holocaust in Fodor’s kitchen. When deciding which information is relevant to a particular task we could not possibly (even using parallel processing) go through an exhaustive list of things that we know, ticking each off as being irrelevant like R\(^D\); rather, we seem to know just what to consider and what not to consider. What seems to be needed for a good cognitive simulation is a system that, while storing a vast amount of information, genuinely ignores most of what it “knows,” and operates with a well-chosen portion of its knowledge at any moment. Thus the frame problem has helped motivate the development of special non-monotonic logics intended to minimize the assumptions that must be retracted given further evidence.

Minimizing assumptions is something that script theory does very well. First, scripts help to provide a context for inference-assumption making, and this is a rather effective way of cutting down the possibilities; it is an effective way of (to use Dennett’s words) “using well-selected elements from [the system’s] store of knowledge about the world it operates in” (1987c: 49). Second, script theory handles non-changes implicitly by allowing the system to assume by default that a state persists, unless there is indication that the state has changed relative to surrounding conditions. Next, scripts permit both expectations and predictions; not just any expectations and prediction, but expectations and predictions that are relevant to the established context: because they are generic knowledge structures, scripts include frames or slots for a variety of different variables, making them able to deal with many different variations on the same theme. This, in turn, produces the

\(^{30}\) All of the citations in this paragraph are found in Pylyshyn’s (1987) anthology, The Robot’s Dilemma: The Frame Problem in Artificial Intelligence.
representations required for drawing situation-specific inferences and making situation-specific forecasts.

Dennett considers script theory, but complains that while script-based systems "perform credibly when the world co-operates with their stereotypes, and even with anticipated variations on them, when their worlds turn perverse, such systems typically cannot recover gracefully from the misanalyses they are led into. In fact, their behaviour in extremis looks for all the world like the preposterously counter-productive activities of insects betrayed by their rigid tropisms and other genetically hard wired behavioral routines" (1987c: 57). This is most likely an exaggeration. We ought to remind ourselves that even very intelligent people make mistakes. We are prone to overlooking retrospectively obvious flaws in our plans and we often fail to make otherwise clear inferences based on well-perceived and understood external data. In any case, there is nothing in Dennett’s criticisms which point to an inherent defect in the script idea. Of course we are much more capable than SAM, for SAM is essentially starting from scratch, while we are the crux of many, many centuries of evolutionary fine-tuning. The problem, I suggest, is not in the idea of a script at all, but in the limitations of the AI programs in which these scripts operate. Dennett seems to acknowledge at least this much when in the very next passage he observes:

When these embarrassing misadventures occur, the system designer can improve the design by adding provisions to deal with the particular cases ... This process of redesign recapitulates the process of natural selection in some regards; it favours minimal, piecemeal, ad hoc redesign which is tantamount to a wager on the likelihood of patterns in future events. So in some regards it is faithful to biological themes. (1987c: 57)

It is the prospect of having to deal with the frame problem that has frightened many philosophers into shunning holism. Gabriel Segal, for example, attacks descriptive theories of narrow content on the grounds that they are radically holistic. He illustrates using the concept of tiger, where again the problem of relevance is referenced:

First, we need to decide which of the various descriptions I associate with tigers is to be counted as relevant. How might we select among, for example, "yellow-brown," "black-striped," "carnivorous," "maneless," "feline," "animate," "furry," "fierce," "apt to growl and roar," and "indigenous to India"? The more superficial descriptions seem more or less on a par. None has a privileged role in determining cognitive content. The kind terms have a greater influence on cognitive role than the superficial ones, since they are embedded in a biological theory and are rich in inductive consequences. The fact that tigers are animals generates more interesting consequences than the fact that they are yellow-brown. Moreover, it is reasonable to regard animacy as a definitive property of tigers. That is why robot tigers are ipso facto not tigers but albino tigers are... The associated kind concepts do not suffice to determine cognitive role, because we know of too many species of carnivorous felines. So the superficial features must be included somehow. But there really seems to be no way to include some and exclude others. So we must include them all ... This account becomes extremely holistic. For theories are not self-contained. It is part of my theory of animals that they are physical objects. But of course physical object is a core concept of another large theory ... Thus to account for a single concept we have to specify the subject's entire theory of the world. (2000:116-17)

We are thus taken from a claim that all of the properties of o must be included in o's description to the conclusion that in order to know what someone’s concept of o is, we have to know that person’s entire theory of the world. Call this the requirement of full conceptual transparency.

The first thing we should say about Segal’s argument is that its conclusion is false: not only do we not have to know very much about the people with whom we carry on a meaningful dialogue,
but in most cases we actually do not possess any such knowledge. Yet we speak in a perfectly coherent and comprehensible manner, and so do our conversational partners, despite the fact that we may be strangers to one another. Explaining why full conceptual transparency is not necessary is one of the chief boasts of script theory: we take for granted that others are able to infer all of the relevant details left out of a verbal exchange, because we implicitly know that they have access to, and can therefore consult, the scripts we explicitly reference as well as those we simply assume or take for granted. So whatever problems holism may have, the requirement of full conceptual transparency is not one of them.

Segal also makes a number of other problematic claims:

First, we need to decide which of the various descriptions I associate with tigers is to be counted as relevant...

We need to decide no such thing! The descriptions that one associates with tigers are defined by one's TIGERDs, which develops through experience and indirect learning. For most of us, the tiger scriptD includes frames related to all kinds of facts about the species—physical characteristics, behavioural patterns, social habits, environment needs, etc. My recognition that o is a tiger will depend on which subset of these frames is stimulated in a given situation and on whether or not the stimulated subset has surpassed the relevant activation threshold for TIGERDs. So what may be a relevant frame (piece of knowledge, bit of description) in one instance may not be so in another. Context makes a difference. But once TIGERDs is triggered and subsequently enables the recognition of o as a tiger, all of the frames of that scriptD become applicable to o. Whether or not the subject will be consciously aware of all of these frames again depends on the nature and context of the situation.

How might we select among, for example, "yellow-brown," "black-striped," "carnivorous," "maneless," "feline," "animate," "furry," "fierce," "apt to growl and roar," and "indigenous to India"? The more superficial descriptions seem more or less on a par. None has a privileged role in determining cognitive content...

Leaving aside the word "superficial" for the moment, I agree that none of the "descriptions" of a tiger has an inherently privileged role in determining the cognitive content of TIGERDs. But I do maintain that certain kinds of "descriptions" are made privileged by the context of the situation in which recognition occurs and the kind of information processed therein. Suppose Samson is hunting in the Indian jungle and spots what looks to him like a tiger about thirty meters ahead of his location. His glimpse was quick and not well studied, but he still decides with confidence that the animal he saw was a tiger. And indeed it was. Samson’s recognition of o as a tiger was in this case based solely on visual cues. Thus within the context of this encounter—and given the mode of information that was then available to him—we can think of those structures in Samson’s brain that have been selected to react to the physical characteristics of tigers as “privileged”—not inherently so, but contextually so.

On the other hand, context can be extended to include all past relevant encounters and learning episodes, in which case we may reason that some of the frames of Samson’s TIGERDs may have become “privileged” simply in virtue of the manner in which he learned about tigers, the manner in which his tiger scriptD was formed. Perhaps Samson’s past learning episodes inadvertently emphasized some kinds of descriptions over others. If Samson nevertheless manages a well-rounded understanding of tigers, this sort of emphasis or bias will not have a great effect on his concept of the animal, but it will probably put the emphasized frames in the “forefront” of his TIGERDs, and he may well think of these as the most obvious or typical characteristics of the species.

See §2.2.ii for a discussion of Intentionality and consciousness.
Moreover, it is reasonable to regard animacy as a definitive property of tigers. That is why robot tigers are ipso facto not tigers but albino tigers are not...

This is fairly uncontroversial, just as long as we understand that object recognition—or better yet, the application of a script to some object—requires no such stringent condition. So while certain facts or descriptions may be indispensable to tigerhood, no frame is indispensable for Samson’s applying his TIGERDS (correctly or incorrectly) to some object.

The descriptions to which Segal refers are here conceived as the different frames which constitute a script. These frames either store data (about the physical and non-physical characteristics of things, situations, goals, and events), or are sensitive to the transmission of certain information by signals in various modes. However, it is not the type of informational signal—whether it is visual, auditory, linguistic, or otherwise—or the nature of the information itself that determines narrow cognitive content; it is the scripts that such signals activate. Fortunately, scripts already come with their frames in place, as determined by the past learning episodes and experiences of the subject. Thus it makes no sense to wonder which of these frames are “essential” and which are “inessential,” which should be “included” and which should be “omitted.” It is not a philosophical argument or some decision procedure that situates or displaces a scriptal frame, but rather what we learn, how we learned it, and the context in which we apply it.

6.5 Summary

The goal of this chapter has been to provide some account of how scripts may be realized physically. To that end, Dretske’s information-processing approach to Intentionality proved very helpful, for it allowed us to define scripts as brain-structures that have acquired the job of selectively indicating the presence of one or another feature of the external world. The functional selectively of these scripts is, in turn, a consequence of their ability to digitalize incoming analogue information.

Next came our idealized hierarchy of physical events, which gives us an inkling of how the multiple nesting of (sub)goal states—prevalent in instrumental and personal scripts—may be structured both physically and behaviourally. Autoassociation grounded scriptal priming, and LASCH was introduced as a hypothesis with the potential of elucidating a large class of phenomena related to scripts and natural language comprehension. The status of cognitive holism was the final issue we examined. It seems to me that Dretske had it basically right when he suggested that holism is the single most important difference between truly sophisticated information-processing systems and systems of more modest abilities.

Throughout, my aim has been for plausibility, not comprehensiveness. I cannot take credit for most of the ideas presented above, but I do think that they can be expanded and combined in innovative ways. Exploring how some of these ways might profit script theory has been the topic of this chapter.
A Script for Intentionality

When a theory has been apprehended logically, there is often a long and serious labour still required in order to feel it: it is necessary to dwell upon it, to thrust out from the mind, one by one, the misleading suggestions of false but more familiar theories, to acquire the kind of intimacy which, in the case of a foreign language, would enable us to think and dream in it, not merely to construct laborious sentences by the help of grammar and dictionary.


How should we think of mental processes? One way is to think of them as purely computational. Computationalism (or strong Al) is the doctrine that all human mental activity is reducible to algorithms and could therefore be implemented, at least in principle, on a computer. Fodor was once a faithful adherent to this doctrine. He spent the late 70s and most of the 80s trying to construct a theory of how the semantic properties of cognitive states can be linked to their syntactic properties so as to make the mind causally efficacious.¹ Using the analogy of a Turing Machine, Fodor observed that “so long as we are thinking of mental processes as purely computational, the bearing of environmental information upon such processes is exhausted by the formal character of whatever the oracles write on the tape. In particular, it doesn’t matter to such processes whether what the oracles write is true” (1980: 65). This was Fodor’s way of promoting methodological solipsism as a viable approach to cognitive science research, but he has never advertised this strategy as exhaustive of the scope of psychology:

... there is room both for a computational psychology—viewed as a theory of formal processes defined over mental representations—and a naturalistic psychology, viewed as a theory of the (presumably causal) relations between representations and the world which fixes the semantic interpretation of the former. (1980: 66)

Fodor called those mental states falling under the domain of computational psychology opaque, and those falling under the domain of naturalistic psychology transparent. Then he remarked,

ontologically, transparent readings are stronger than opaque ones; for example, the former license existential inferences which the latter do not. But psychologically, opaque readings are stronger than transparent ones; they tell us more about the character of the mental causes of behaviour. The representational theory of mind explains this anomaly. Opaque descriptions are true in virtue of the way that the agent represents the objects of his wants (intentions, beliefs, etc.) to himself. And, by assumption, such representations function in the causation of the behaviors that the agent produces ... My claim has been that, in doing our psychology, we want to attribute mental states fully opaque because it's the fully opaque reading which tells us what the agent has in mind, and it's what the agent has in mind that causes his behavior. (1980: 66-7)

¹ See especially Psychosemantics.
Computationalism has its opponents, of course. One zealous critic is John Searle, whose famous Chinese Room argument (1980) was designed to deal a knockdown blow to strong AI.²

Though the debate regarding the feasibility of computationalism is important, I am more interested in some of the remarks Fodor made in defending this position—particularly, those pertaining to the agent's point of view. There is a lot to be said for understanding the perspective of someone who acts. Such an understanding is necessary, as Fodor maintained, to an account of the person's reasons for behaving in a certain manner. If you want to know why Samson boxed Adam's ears, try to find out something about how Samson interpreted the relevant circumstances. It does not matter whether or not Samson's interpretation is objectively sound, whether or not events unfolded as he believes. But if the object is to get at Samson's motives for behaving as he did, how he represented these circumstances to himself is important. Similarly, if you want to know why Samson reacts in such-and-such way to object o, try to find out something about how he represents o to himself, then you might discover why he eats o, or uses o to scrub his back, or takes o out to dinner and buys it gifts on its birthday.

Samson's perspective is a pervasive part of his psychological life. It gives rise to the dimension of Intentionality we have called directedness. Intentional directedness, along with the closely related aspectual (or perspectival) shape, capture the much-discussed narrowness of Samson's mental states. Aboutness respects the fact that Samson doesn't just spontaneously or randomly represent things, but usually (although not always) does so in causal response to specific external provocations, while error (or misrepresentation) seems to be a natural and expected feature of any information-processing system.

Any plausible account of Intentional content, I have claimed, must explain these four characteristics. What I want to do here is, first, examine how the theory of scripts developed in the previous chapters stands up to our demands for an adequate account of Intentionality, and, second, consider the consequences of the scriptal approach for the various externalist perspectives outlined in chapter III.

7.1 Aboutness, Directedness, and Aspectual Shape

Not all mental states are Intentional,³ and not all Intentional states make use of scripts. My belief that 2 + 2 = 4 does not seem to implicate knowledge that is scriptally organized, at least not in the form of any script-type that we have previously examined. Our scriptal theory of Intentionality, therefore, should be viewed as limited to that subclass of mental states that encompass representations of physical and fictional objects, generic event sequences, personal motives, goals,

---

² As is customary with articles appearing in Behavioral and Brain Sciences, Searle's paper was followed by an open peer commentary. Among the most interesting replies are "Understanding Searle" by Roger Schank, "What Intuitions About Homunculi Don't Show" by Ned Block, "The Milk of Human Intentionality" by Daniel Dennett, and especially "Programs, Causal Powers, and Intentionality" by John Haugeland. Reactions ranged from the decidedly positive to the overtly caustic. Hence B. Libet observes that "Searle shows, in a masterful and convincing manner, that the behavior of the appropriately programmed computer could transpire in the absence of a cognitive mental state" (1980: 434), while Douglas Hofstadter states in his opening sentence, "This religious diatribe against AI, masquerading as a serious scientific argument, is one of the wrongest, most infuriating articles I have ever read in my life" (1980: 433).

³ See §2.1.i.
plans, and schemes. These categories correspond to the kinds of scripts we introduced in chapter IV. There could be more, but we shall content ourselves with what we have.⁴

1. Intentional Aboutness

We saw in chapter II that Intentional states have reference as well as sense, and near the end of chapter VI the same was shown to be true of scripts. Thus, insofar as an Intentional mental state involves one or another script, that state’s reference and sense will be derived from the reference and sense of the script in question. Reference has been identified as Intentional aboutness and sense as Intentional directedness. In order to find out what an Intentional state may be about and to what it may be directed, it was suggested that we construct for each Intentional state an aboutness extension (Eₐ) and a directedness extension (Eₙ).

What goes into the Eₐ of an Intentional state is whatever real or actual object bearing C to the representation of the agent whose Intentional state it is. C is a causal relation, in consequence of which a representation is primed. The primed representation is a particular script which is a product of a system’s ability to digitalize incoming analogue information. The aboutness of an Intentional state, then, may be defined in terms of a causal relation connecting a real-world object with a primed script in an agent. The connection need not be direct, but the aetiology of C must trace back to a real object in order for the relevant Intentional state to have aboutness.⁵

There is nothing new about positing an agent-environment causal relation as a way of grounding Intentional content. But the criterion that aboutness must connect an Intentional state with a real aspect of the real world is (as far as I can tell) new. This condition divides a person’s repertoire of scripts into two rough categories—those that have aboutness (α), and those that don’t (φ). To the first kind belong scripts like CHAIR₉₅, RESTAURANT₉₅, ROBBERY₉₅, DRIVING-A-CAR₉₅, and MY-GRANDMA₉₅. We shall impose neither temporal nor spatial stipulations for membership in α, which is to say that a script with aboutness need not represent the current world, nor must the represented(s) be physically near the agent. There are no spatiotemporal conditions for membership in α because aboutness is concerned only with whether or not the object referenced by a script is actual, not whether or not it is proximate or current. Thus we also include in the α-category scripts like CHARLES-DICKENS₉₅ and TRICERATOPS₉₅.

One of the advantages of LASCH⁶ is its demystification of how we can form, by means of indirect learning, scripts for things we have never experienced before. And not just any scripts, but

⁴ Unless otherwise noted, then, any future reference to “Intentional states” will refer to those states that putatively have underlying scripts.

⁵ Kripke gives an example where a chain of communication establishes C. “Someone, let’s say, a baby, is born; his parents call him by a certain name. They talk about him to their friends. Other people meet him. Through various sorts of talk the name is spread from link to link as if by a chain. A speaker who is on the far end of this chain, who has heard about, say Richard Feynman, in the marketplace or elsewhere, may be referring to Richard Feynman even though he can’t remember from whom he first heard of Feynman or from whom he ever heard of Feynman. He knows that Feynman is a famous physicist. A certain passage of communication reaching ultimately to the man himself does reach the speaker. He then is referring to Feynman even though he can’t identify him uniquely. He doesn’t know what a Feynman diagram is, he doesn’t know what the Feynman theory of pair production and annihilation is. Not only that: he’d have trouble distinguishing between Gell-Mann and Feynman. So he doesn’t have to know these things, but, instead, a chain of communication going back to Feynman himself has been established, by virtue of his membership in a community which passed the name on from link to link ... ” (1980: 91).

⁶ See §6.3.ii.
scripts that are elaborate, detailed, and fully capable of intermingling with other concepts and finding a place in an individual's network of ideas:

*If language developed as an alternate informational capacity, then we can learn about things—we can form all kinds of scripts—without necessarily having direct sensual contact with the things learnt. We can do this because natural language is not confined, as the five senses usually are, to what is true of the world or spatiotemporally close.*

To the θ-category belong scripts like **SANTA-CLAUS**, **DAVID-COPPERFIELD**, and **PAPA-SMURF**. These scripts are not about anything, but they obviously still have content. The reason mental states need not be about an actual thing in order to be Intentional has to do with the way scripts carry out their representational work: all that is required for a script to have content—and so manifest Intentionality—is that it represent a *specifiable* o, not necessarily an *actual* o. The frames of the script, in other words, must pick out well-defined properties and use these properties in the process of identification and recognition, regardless of whether or not the properties so picked out "adhere" to a tangible thing. It is fairly obvious, I think, that Santa Claus, David Copperfield, and Papa Smurf all have such well-defined, specifiable properties.

Notice that the mechanisms which enable θ-category scripts are the very ones that make indirect learning possible. In both cases, LASCH hints at an explanation:

*LASCH explains imagination, our ability to create fictional people, places, and events. Because we do not have to actually experience objects or situations through the senses in order to build scripts of them, we can make a Santa Claus and a Tooth Fairy. These "creatures" get definitional scripts—complete with physical, social, and behavioural characteristics—just as all other objects do. The only difference is the mode in which the relevant data is conveyed, plus the fact that the ontological status of these creations is known to most people. Imagination is thus an extension of indirect learning, for similar enabling processes underlie both phenomena.*

It is not unreasonable, I suppose, to wonder how scripts came to have aboutness in the first place, though the question is not terribly difficult to answer. A straightforward evolutionary story will do the trick: faithful, detailed representations of the environment are highly adaptive for organisms, and the better an organism is at representing external conditions the more likely it is to find food and avoid becoming some, to search for shelter and cope with changing ecosystems, to fight off enemies and find a mate or two. We might *still* have done well if we, like most other organisms, were able only to represent the world in a very basic and limited fashion. But humans do more than just represent; humans hyperrepresent, and *labelling* is the key. We took these simple object-scripts and gave them headers, which initiated a huge "conceptual expansion campaign" in which we relentlessly mixed, matched, and connected scripts into sophisticated cognitive networks.

Though I favour an evolutionary account of the origin of aboutness, I am wary of teleological talk. From the time of Charles Darwin, evolutionary biologists have insisted that there is no goal-directedness in the mechanisms by which genetic mutations are favoured under particular survival pressures. And indeed, if we accept natural selection as a process based on *random* genetic mutation, it is very difficult to see just how purpose or design might accompany whatever long-term changes a species might manifest as a result of such a process.

But aside from these considerations, teleology has always seemed to me to be one step up—and not a very big step, at that—from the occult: it is a short road from a purposive conception of
nature to the kind of medieval mindset that saw spirits haunting forests and the wickedness of man punished by hailstorms and floods. A simple, non-teleological evolutionary story of aboutness suites me better. Such an account would also mitigate worries about solipsism: there is no question of an outside world, since scripts are formed first and foremost for the competitive advantage of an organism, and so presuppose the external world and all its "belongings." Very little about script theory would make sense otherwise.7

ii. Intentional Directedness

What goes into the Ej of an Intentional state is any object—real or non-real—that answers to the frames of the script underlying the state. Let's call the non-modal reading one that limits "any" to actual objects, and the modal reading one that extends "any" to include possible but non-existent items. It is not enough to conceive of directedness on the non-modal reading, because that excludes aboutless Intentional states involving fictions or fabrications. The modal reading, in contrast, is more encompassing; it allows us to have a unified treatment of the directedness of all Intentional states, those pertaining to real objects and those pertaining to fictional entities.8 What I mean by an object answering to the frames of a script is simply the frames' reacting causally to the object: the neural structure constituting the script would respond to the object if it were present. Hence, any Intentional state has two extensions, Ea and Ed, and Ea might not be equivalent to Ed.

Directedness (unlike aboutness) is sensitive to what a system can and cannot discriminate. Hence, Dretske claims that whether or not two concepts are identical is "not simply a matter of knowing, or not knowing, the right labels or words for experienced differences. It is, instead, a matter of lacking [or having the relevant] discriminatory powers" (1995: 138). Fodor makes a similar point when he comments that if "a creature can't distinguish Xs from Ys, it follows that the creature can't have a concept that applies to X but not to Y" (1994: 32). I have tried to articulate the same intuition in the previous chapter using the idea of informational content:

A loss in informational content produces a loss in the discriminatory powers of an information-processing system, and the effect is proportional: the more data lost, the more diluted a system's ability to discriminate becomes.

This can be expressed more precisely as the principle of Intentional inclusion (PII).

7 Attribution theorists in psychology have tried to specify why solipsism is not a psychologically real possibility for most people. The following from H. H. Kelley addresses this point:

When a person has an impression that something is true of an entity, how does he ascertain that the impression reflects the inherent properties of the entity and not his own characteristics ...? The four criteria for external validity are ...

1. Distinctiveness: the impression is attributed to the thing if it uniquely occurs when the thing is present and does not occur in its absence.
2. Consistency over time: each time the thing is present, the individual's reaction must be the same or nearly so.
3. Consistency over modality: his reaction must be consistent even though his mode of interaction with the thing varies. (For example, he sees it to have an irregular outline and he feels it to be rough; or first he estimates the answer to the problem and then he calculates it.)
4. Consensus: attributes of external origin are experienced the same way by all observers.

To the degree a person's attributions fulfill these criteria, he feels confident that he has a true picture of his external world. He makes judgements quickly and with subjective confidence, and he takes action with speed and vigour. When his attributions do not satisfy these criteria, he is uncertain in his views and hesitant in action. It is not assumed that fulfillment of these criteria implies veridicality of the person's attributions. The specific evidence provides a basis for subjective validity (as manifested, for example, in confidence in the validity of one's attributions) but not necessarily a basis for their objective validity. (1967: 197-8.)

8 My thanks to Dr. Paul Bartha for helping me sort out these points.
PII: An information-processing system (agent) A cannot be in a script-based Intentional state \( I_s \) directed exclusively toward an actual or possible object \( o \) if \( A \) is incapable of discriminating between \( o \) and other superficially similar actual or possible objects, \( o_2, o_3, o_4, \ldots \).

The reasoning behind PII is that, since the object(s) which go into the \( E_d \) of an Intentional state are determined by a script, that Intentional state does not identify in those objects any differences that cannot be identified by the script itself. In other words, where \( A \) is an information-processing system, \( I_s \) a script-based Intentional state of \( A \), and \( E_d(I_s) \) the directedness extension of \( I_s \), PII essentially limits the individuation of all \( o \in E_d(I_s) \) to the range of discriminations made by \( A \) based on \( A \)'s activated script.

In order to say exactly what a directedness extension is, we shall use the intuition behind PII to construct an equivalence class of actual and possible objects relative to an arbitrary definitional script \( S \). An equivalence class is determined by means of an equivalence relation. Mathematically, a relation \( R \) on a given domain \( \Omega \) is an equivalence relation on \( \Omega \) if and only if (iff) \( R \) is reflexive on \( \Omega \), \( R \) is symmetric on \( \Omega \), and \( R \) is transitive on \( \Omega \). \( R \) is reflexive on \( \Omega \) iff for all \( o_i \in \Omega \), \( o_i R o_i \). \( R \) is symmetric on \( \Omega \) iff for all \( o_i \) and \( o_j \in \Omega \), if \( o_i R o_j \), then \( o_j R o_i \). \( R \) is transitive on \( \Omega \) iff for all \( o_i, o_j, \) and \( o_k \in \Omega \), if \( o_i R o_j \) and \( o_j R o_k \), then \( o_i R o_k \).

Let \( S \) be an arbitrary script \( S \) and \( F = \{ f_1, f_2, f_3, \ldots, f_n \} \) be the set of frames constituting \( S \). For any two actual or possible objects \( o_i \) and \( o_j \), we say that \( o_i R o_j \) iff for every \( f_i \in F \), both \( o_i \) and \( o_j \) possess the feature or characteristic or property to which \( f_i \) is selectively sensitive. Now we let \( \Omega \) be the set of all actual or possible objects for which the relation \( R \) holds.

The first thing to point out is that \( \Omega \) is a well-defined set. Well-definedness neither entails that the set be finite nor that its members be corporeal objects. Rather, what well-definedness requires is that there be a decision procedure, a definite "yes" or "no" answer to the question whether the item being contemplated belongs in the set. That is precisely what we have for \( \Omega \): any actual or possible object \( o_i \) will be a member of \( \Omega \) iff for every \( f_i \in F \), \( o_i \) possesses the feature for which \( f_i \) has acquired the job of indicating. Otherwise, \( o_i \notin \Omega \).

Second, it is not difficult to see that \( R \) is an equivalence relation on \( \Omega \): for every \( o_i \in \Omega \), \( o_i \) has the same properties as itself (including those properties to which all \( f_i \in F \) are responsive); so \( o_i R o_i \) is true (reflexivity). For any two objects \( o_i \) and \( o_j \in \Omega \), if \( o_i \) has the same properties picked out by all \( f_i \in F \) as does \( o_j \), then \( o_j \) also has those properties in common with \( o_i \); so \( (o_i R o_j) \rightarrow (o_j R o_i) \) is true (symmetry). For any three objects \( o_i, o_j, \) and \( o_k \in \Omega \), if \( o_i \) has the relevant properties in common with \( o_j \) and \( o_j \) has those properties in common with \( o_k \), then \( o_i \) also shares the relevant properties with \( o_k \); so \( [(o_i R o_j) \& (o_j R o_k)] \rightarrow (o_i R o_k) \) is also true (transitivity).

Now we define the directedness of an Intentional state using \( R \) as follows:

Let \( I \) be any Intentional state, \( S \) be the script \( S \) underlying \( I \), \( F = \{ f_1, f_2, f_3, \ldots, f_n \} \) be the set of frames constituting \( S \), and \( o \) be any actual or possible object such that, for every \( f_i \in F \), \( o \) displays the feature for which \( f_i \) is selectively sensitive. The equivalence class of \( o \) as determined by \( R \) is fixed by the set \( o_i/R = \{ o_j \in \Omega \mid o_i R o_j \} \).

---

9 This is read "the class of \( o_i \) modulo \( R \)" or simply "\( o_i \mod R \)"
In plain English, what $o_i \mathcal{R}$ does is this: it picks out, for any Intentional state involving a definitional script, all actual and possible objects that possess the total sum of features or characteristics which the frames of the script have the job of indicating. Hence, all of these objects will be equivalent or identical or indistinguishable (for the agent whose script it is) relative to the script in question. If we let $"I_s"$ designate the appropriate script-based Intentional state, then $E_d(I_s) = o_i \mathcal{R} = \{o_i \in \Omega \mid o_i \mathcal{R} o_j\}$.\(^{10}\)

One final observation. The way $\mathcal{R}$ is defined does not commit us to specifying the kind of features, or characteristics, or properties to which the frames of our arbitrary script may be sensitive. What $\mathcal{R}$ does is define an equivalence class relative to the manner in which the frames of some script represent, regardless of what actual properties are being represented by these frames. The external properties may be whatever you like; they may even be radically different from one object in the equivalence class to another. The important thing is that they must elicit the same reaction from the relevant definitional script, in which case the script would represent all of the corresponding objects as identical whether or not they are truly so. That is what it means to have an equivalence class relative to a definitional script: the equivalence of the members comes from the way in which their associated properties affect the script, not from the sort of properties they in fact have. Without relativity to a script, $\mathcal{R}$ would not be an equivalence relation.\(^{11}\)

**Designatory and Indicatory Rigidity**  
Our definition of $E_d(I_s)$ makes it obvious that, as far as directedness is concerned, scripts are not rigid indicators.\(^{12}\) Recall Kripke’s use of “rigid designators” in *Naming and Necessity*. First, Kripke outlined what a proper conception of “possible world” should be:

An analogy from school... will help to clarify my view. Two ordinary dice... are thrown, displaying two numbers face up. The thirty-six possible states of the dice are literally thirty-six ‘possible worlds’... [T]he ‘actual world’ in this case is the state of the dice that is actually realized ... But when we talk in school of thirty-six possibilities, in no way do we need to posit that there are some thirty-five other entities, existing in some never-never land, corresponding to the physical object before me ... ‘Possible worlds’ are little more than the miniworlds of school

\(^{10}\) The possibility of defining $E_d(I_s)$ using the notion of an equivalence class occurred to me independently. But I was very gratified when, during the course of my research, I came across three references—aside from Dretske (1995: 139) and Fodor (1994: 32)—to the same basic idea. The first was made by Christopher Peacocke in his (1996) essay concerning externalism and self-knowledge. Peacocke imagined a critic defining a relation of “rational similarity” between two concepts and then using that relation to work out the discriminatory limits of narrow contentful states. The second reference, ironically, was made by Tyler Burge in the context of his (1998) discussion of “Disjoint Type” concepts and “Amalgam Type” concepts. Disjoint Types occur when environmental changes (e.g., slow switching between Earth and Twin-Earth) produces similar but unique concepts in an individual (e.g., aluminium and twaaluminium); Amalgam Types occur when, because of an impoverished ability to discriminate certain things, an individual develops a broadened concept that envelops both the native object and its counterpart on Twin-Earth. The third reference was made by Gabriel Segal (2000), who assigned “motleys” (i.e., assortments of objects falling under one or another description) to the extensions of some Intentional states. Neither one of these authors specifically uses equivalence classes, but the hints are nevertheless there.

\(^{11}\) For a concrete instance of a definitional script and the equivalence class to which it gives rise, we might refer to eighteenth-century Adam’s script of water. Adam’s script lacks any data about the chemistry and molecular composition of water, but it contains frames related to observable physical characteristics, countless functional properties, and many other mundane particulars. But, by assumption, none of these frames, either individually or in combination, could distinguish $\text{H}_2\text{O}$ from $\text{XYZ}$. Furthermore, there are numerous other compounds—some real, many imaginary—that could easily pass the ‘water test’ as determined by Adam’s definitional script. It is these compounds that constitute the equivalence class to which Adam’s $\text{WATER}_{\text{DS}}$ gives rise.

\(^{12}\) At least not normally so. In order for a script to indicate rigidly, it must satisfy what I shall call *bi-extensional equivalence*. I return to this point in §7.2.i.
probability blown large ... [They] are total 'ways the world might have been', or states or histories of the entire world ... [In practice we cannot describe a complete counterfactual course of events and have no need to do so. A practical description of the extent to which the 'counterfactual situation' differs in the relevant way from the actual facts is sufficient; the 'counterfactual situation' could be thought of as a miniworld or a ministate, restricted to features of the world relevant to the problem at hand ... There is nothing wrong in principle with taking these [possible worlds], for philosophical or for technical purposes, as [abstract] entities—the innocence of the grammar school analogue should allay any anxieties on that score. (1980: 16-19)

Kripke next tells us what rigid and accidental designators are: “Let’s call something a rigid designator if in every possible world it designates the same object, a non-rigid or accidental designator if that is not the case ... One of the intuitive theses I will maintain in these talks is that names are rigid designators. Certainly they seem to satisfy the intuitive test [definition] mentioned above: although someone other than the U.S. President in 1970 might have been the U.S. president in 1970 (e.g., Humphrey might have), no one other than Nixon might have been Nixon. In the same way, a designator rigidly designates a certain object if it designates that object wherever the object exists” (1980: 48-9). Like names, natural-kind terms are also rigid designators:

Even though we don't know the internal structure of tigers, we suppose—and let us suppose that we are right—that tigers form a certain species or natural kind ... we can say in advance that we use the term ‘tiger’ to designate a species, and that anything not of this species, even though it looks like a tiger, is not in fact a tiger. (Kripke 1980: 120-1)

Now comes the point: unlike names and natural-kind terms, scripts do not designate things; they indicate things or aspects (maybe real, maybe invented) of the external world. Thus the question of whether or not scripts accomplish their indicatory duties in a rigid way can only be answered by evaluating the discriminatory power of an information-processing system $A$. If it is found that $A$ cannot discriminate between $o_i$ and $o_j$ given the totality of its resources at time $t$, then $A$ cannot rigidly indicate, at time $t$, the presence of either $o_i$ or $o_j$. This is the chief claim of PII.

Kripke explained that names and natural-kind terms can do their job rigidly because the act of designation does not depend on the power or ability of the $A$ that uses designating terms to know anything particular about the designata of these terms. The act of indicating, in contrast, does depend on the ability of $A$ to discriminate between the things indicated. We can, of course, insist that $A$ does indicate the presence of both $o_i$ and $o_j$, if we understand “indicate” to mean something like “react to in a specific way.” But from the perspective of $A$ itself, there can be no difference between $o_i$ and $o_j$ unless the system has some way of representing this difference. Where the relevant distinctions are

13 Robert Stalnaker (1975: 66-7) makes use of a similar notion of possible worlds: “It is a common and essential feature of such activities as inquiring, deliberating, exchanging information, predicting the future, giving advice, debating, negotiating, explaining and justifying behavior, that the participants in the activities seek to distinguish, in one way or another, among alternative situations that may arise, or might have arisen. Possible worlds theory, as an explanatory theory of rational activity, begins with the notion of an alternative way that things may be or might have been (which is all that a possible world is) not because it takes this notion to be unproblematic, but because it takes it to be fundamental to the different activities that a theory of rationality seeks to characterize and relate to each other. The notion will get its content, not from any direct answer to the question, what is a possible world? or from any reduction of that notion to something more basic or familiar, but from its role in the explanations of such a theory. Thus it may be that the best philosophical defense that one can give for possible worlds is to use them in the development of substantive theory.”

14 But perhaps $A$ can rigidly indicate the whole equivalence class of $\mathcal{R}$—that is, the totality of actual or possible objects which display all the characteristics to which the frames of $O_\mathcal{R}$ are sensitive. Be that as it may, we are here talking about the members of the equivalence class, not the class itself. $A$ cannot rigidly (i.e., exclusively) indicate $o_i$ or $o_j$, but it may very well be able to rigidly indicate $o_\mathcal{R}$.
lacking, *A* can never really "know"—based on its resources alone—whether it is indicating the presence of *o₁* or the presence of *o₂*.

If this is correct, then there is a very real disparity between scripts and at least some designating expressions. This, I want to suggest, is symptomatic of a more general cleavage between *linguistic meaning* and the content of *narrow Intentional states*. To be sure, we use language to express what we think all the time, and we do so effectively and without much trouble. Nevertheless, there *is* a difference between what we think and the words we choose to express our thoughts. This difference has to do with how the semantic properties of words and contentful Intentional states are "anchored." We may follow Burge, Kripke and others in supposing that the meaning of linguistic expressions is anchored *externally*, by particular rules of usage or social conventions at a given time. Narrow content, on the other hand, is anchored *internally*; it depends on the kind of information the system can process, the manner in which it represents this information, and the range of discriminations made by the relevant representations.¹⁵

**Fictional Objects** We constructed our definition of *E_d*(*Lₐ*) specifically to deal with real and non-real objects. Thus *o₁*, if *R* picks out *Santa Claus* and everything *Santa*-like if *S* were *SANTA-CLAU$S_{₁ₙ}$*, or *PETER PAN* and everything *Pan*-like if *S* were *PETER-PAN₂ₙ*. The motivation behind doing things in this way is to have a unified treatment of Intentional directedness irrespective of whether or not scriptal representations have aboutness. This is in keeping with what I have suggested back in §2.2.iii, which is to conceive of the object(s) of directedness *counterfactually*, not as items that do satisfy the definitional script in question, but as any item that *would* satisfy the script if it possessed all the relevant features. I think this strategy yields good results, especially if we are careful to heed Kripke’s warning and avoid thinking of our innocent fictions as Meinongian artefacts that exist in some "never-never land."¹⁶

What I propose we do is to think of imaginary objects as *abstract entities* whose properties are defined by certain definitional scripts, and we can use the headers of these scripts to designate and talk about those entities: hence "Cinderella," "Yeti," and "Loch Ness Monster." We don’t even have to call them "abstract" entities, if the word "abstract" perturbs the philosophical nerves. Choose whatever comforting name you like, only be aware that, so far as the manner in which we learn about the characteristics of these fictions is concerned, there is little difference between them and objects that once existed but are no longer here.¹⁷

Giving fictitious entities names or titles is nothing more than a convenient way to effect quick and easy reference. Thus we are entitled, hopefully without causing confusion, to construct an *E_d* of Eva’s belief that Leprechauns exist by defining an equivalence class of *o₁* as *o₁* determined by her *LEPRECHAUNₙ*. And if we get complaints that such an equivalence class would never get off the ground because there are no Leprechauns—because the set \( \{ l | l \text{ is a Leprechaun} \} = \emptyset \)—we can gently but firmly point out to the complainant that the objection is based on conflating aboutness and

---

¹⁵ I think here is where Segal’s proposal for the use of neologisms might prove helpful (see §3.5).

¹⁶ Kripke was not the first to sound this particular alarm. Russell did it more than half a century prior: “Take, e.g. ‘the golden mountain does not exist’ or ‘the round square is self-contradictory’. If we are to preserve the duality of meaning and denotation, we have to say, with Meinong, that there are such objects as the golden mountain and the round square, although these objects do not have being. We even have to admit that the existent round square is existent, but does not exist. Meinong does not regard this as a contradiction, but I fail to see that it is not one” (1917: 162).

¹⁷ That is not to say that there is no difference absolutely, especially (as we shall shortly see) when it comes to error.
directedness, that the latter is neutral with respect to what is and isn’t actual, and that it is impossible to have an empty directedness extension of any Intentional state, since that would mean that the agent whose state it is has no perspective whatsoever, and so there is no Intentional state to begin with.

**Privileged Access** Recall the controversy about the privileged access one supposedly has to his or her personal Intentional states. There seems to be a fairly wide consensus that such first-person authority is a real phenomenon. The debate has to do with whether or not this authority is threatened by externalism.

My own position is that privileged access is incompatible with externalism. What one often finds in the compatibilist literature are variations on the following retort: “Yes, perhaps Putnam’s Twin-Earth does show that there is a sense in which a person may not know what he is thinking about, but that is just a version of a more familiar sceptical argument. The correct method for answering this kind of scepticism is to insist that thinkers can distinguish the actual world from other *relevant* alternative states of the world; and the possibility that XYZ rather than H₂O flows in rivers is just not a relevant alternative.” This, for instance, is the reply Stalnaker gives in “Narrow Content.” But while I agree that this is a sound strategy for deflecting scepticism, there is a problem of consistency in using it to answer those who do not believe that externalism can be reconciled with first-person authority. We cannot simultaneously agree to accept Twin-Earth scenarios “for the sake of argument,” use them to further externalist intuitions, then simply dismiss them as irrelevant when it comes to other matters. If we are willing to accept an externalist psychology based on Twin-Earth thought-experiments, then we must also be willing to accept whatever other conclusions the same thought-experiments commit us to. In short, what makes the possibility of an XYZ-containing world *relevant* is the fact that we have agreed to take seriously certain counterfactuals involving Earth and Earth₂, along with conditions like slow-switching between the two worlds.¹⁸ So either we get rid of Twin-Earth or we keep it, but the externalist can’t have it both ways.

But let us leave that aside for now. Assuming that the phenomenon is legitimate, what does the scriptal approach to Intentionality make of it? The short answer is that script theory preserves first-person authority, but in a qualified form.

In §3.2.vii, the issue was rephrased in terms of an *asymmetry of epistemic access* between personal and non-personal cognition. The claim was that in the personal case, no empirical evidence is required for knowing that one is having such-and-such thought, whereas empirical evidence is always necessary in the case of others. This manner of putting things makes our way to tackling privileged access clear, if not simple or routine. It is obvious, for instance, that an agent can have no authoritative knowledge of the *aboutness* of his or her Intentional states; aboutness is a causal connection that often has to be empirically investigated in order to be discovered. So the possibility of direct, non-empirical knowledge, if it is true at all, must be true of Intentional *directedness*.

Suppose, then, that Delilah’s Oₜ is not informationally rich enough to discriminate between o₁, o₂, and o₃, all of which are real objects in the world. What implications does this have for the problem at hand? Well, Delilah’s lack of discrimination means that her Eₜ(L) will be inclusive of the set {o₁, o₂, o₃}. Now, how can Delilah know whether her thought is directed toward o₁, o₂, or o₃? This might seem like a legitimate question until we notice just what we are asking. We are *not* asking about Delilah’s point of view; if we were, then the correct answer should be plain: Delilah’s

Intentional state is directed *equally* toward \( o_i, o_j, \) and \( o_k \), because it makes no distinctions between these objects. Furthermore, no matter how intimately or profoundly Delilah might be sensible of her Intentional state, she would still be unable to make the relevant distinctions. So making these distinctions is not indicative of the extent to which Delilah might know (the directedness of) her own thought.

If that seems unsatisfactory, it is because we are really asking a *different* question. We want to know how Delilah can find out what her Intentional state picks out in the world. If this is the case, then we have just stepped out of bounds: Delilah does have direct access to her perspective, which, as it happens, does not distinguish \( o_i, o_j, \) and \( o_k \). Ask her anything about her state of mind, short of questions that presuppose the distinctions in question, and she will answer definitively and without having to investigate her answers empirically. But if you ask what objects her Intentional state picks out in the world—whether it is \( o_i, o_j, \) or \( o_k \)—then you slip into the domain of aboutness, and we have already decided that Delilah has no special privilege in determining what her thoughts might be about. The point is that her inability to distinguish \( o_i, o_j, \) and \( o_k \) does not mean she is oblivious to what is going on in her mind. On the contrary, the fact that she cannot tell you which \( o \) she is thinking about perfectly reflects her \( ED(\omega) \), since her \( OD(\omega) \) in fact does not make the required *discriminations*. The same considerations apply even if \( o_i, o_j, \) and \( o_k \) were fictitious.

This analysis preserves the distinction between the asymmetry of epistemic access to first- and third-person Intentionality, and the completely different notion of *infallible* self-knowledge. The term "asymmetric" forces us to acknowledge *only* that the way we normally come to know our own thoughts is different from the way we normally come to know the thoughts of others. But it in no way forces us to hold that such knowledge must be error-free. Even though Delilah has direct knowledge of her narrow contentful Intentional states, she may still have something to learn—namely, that she cannot distinguish between a certain set of objects, and consequently, that her beliefs may not be carving up the world as well as she supposed. That is about the best we can get out of a notion of first-person privilege that does not entail infallibility.

### iii. Aspectual Shape

I previously stressed that aboutness and directedness are such that neither determines the other. That is to say that given a script-based Intentional state \( I_s \), \( ED(I_s) \) does not determine \( ED(\omega) \) and \( ED(\omega) \) does not determine \( ED(I_s) \). Fregean cases—e.g., Oedipus and the Morning-Evening Star—tell us why aboutness does not determine directedness: it is possible to have Intentional states with identical aboutness but distinct directedness. Putnam's Twin-Earth story tells us why directedness does not determine aboutness: it is possible to have Intentional states directed at the same object(s), but about different things. Scripts make sense of both of these facts. Let us start with Twin-Earth.

**Twin-Earth Cases** According to PII, an information-processing system cannot be in an Intentional state directed exclusively toward \( o \), if it is incapable of discriminating \( o \) from similar objects. This principle served as the basis for a definition of the \( ED(\omega) \) of an Intentional state using the mathematical notion of an equivalence class. From that definition we get the following result:

Both Adam and Adamte's belief that there is "water" in the pitchers before them is directed as much toward \( H_2O \) as it is toward \( XYZ \)—provided that (a) their \( WATER_{\omega} \) is the same, and (b) their \( WATER_{\omega} \) does not distinguish between \( H_2O \) and \( XYZ \).
By premise, the experiential histories of Adam and his doppelgänger are the same: they have witnessed, participated in, and been shaped by type-identical events at the same relative time; they share the same behavioural proclivities, the same inner causal states, the same qualitative experiences; they possess the same mental constructs and share the same language; finally, they have always been physiological duplicates. We therefore have every reason to believe that both individuals, because of their parallel experiential and episodic lives, have acquired the same capacity for analogue-to-digital data transformation performed by physically identical neuronal structures in their brains. In short, the twins have acquired the same definitional scripts of objects with which they have come in contact—including, of course, perceptually indistinguishable items like H\textsubscript{2}O and XYZ.

Script theory thus explains something that has utterly baffled Tyler Burge:

No English–to–English dictionary would give ‘water’ as the entry for the Twin-Earthian’s word. It would thus be a mystery how a Twin-Earthian could share any of Adam’s attitudes that involve the notion of water. They have not had any of the normal means of acquiring the concept. The correct view is that they have acquired, by entirely normal means, a concept expressed in their language that bears some striking, superficial similarities to ours. (1982:110)

Burge’s mistake is twofold. First of all, it is not the relevant concepts (i.e., scripts\textsubscript{D}) operating in the twins that bear a superficial similarity to one another; these concepts are the same. Where there is a superficial similarity is between the objects (H\textsubscript{2}O and XYZ) that the concepts are supposed to represent, and it is the macrophysical similarity of these objects that is behind the identity of the concepts. Secondly, the twins have indeed had all of the normal means of acquiring their joint concept of “water.” Adam, for instance, has experienced all the usual ways of attaining a concept of XYZ, even though he has never had any causal interaction with XYZ. How can this be? Simply put, Adam had the requisite connection to H\textsubscript{2}O, and this allowed him to form exactly the same script\textsubscript{D} of “water” as he would have formed had it been XYZ with which he interacted. What neither of the twins was exposed to, however, was a way—any way—of distinguishing between the two liquids. The upshot is that neither H\textsubscript{2}O nor XYZ was necessary for the formation of the twins’ joint WATER\textsubscript{D}, but each was individually sufficient to develop such a construct. That is how Adam was able to form a concept of XYZ even though he has never had physical contact with the stuff.\textsuperscript{20}

Part of the problem, I think, is that Burge’s attitude presupposes that concepts, like names and natural-kind terms, are rigid. This is why he starts his observation by saying that “no English–to–English dictionary would give ‘water’ as the entry for the Twin-Earthian’s word.” But script-based concepts do not indicate the presence of objects in a rigid manner, and so there is no “mystery” regarding how Adam, who has had no causal interaction at all with XYZ, could have acquired the concept of water\textsubscript{E}. The answer is that his script\textsubscript{D} of water, by virtue of its plasticity and the manner in which it was formed, already is directed toward XYZ.

Adam and his twin are able to form beliefs with respect to the stuff they call “water” because they have learned to extract certain salient features of the liquids—e.g., wetness, colourlessness, odourlessness, tastelessness, along with tactile and functional properties. When the twins encounter objects that manifest these properties, their script\textsubscript{D} of “water” is activated. Not only are these definitional scripts identical, but the conditions under which they have been historically primed are also identical. The twins’ beliefs, therefore, have the same E\textsubscript{d}(L). But having the same E\textsubscript{d}(L)—i.e.,

\textsuperscript{19} Well, almost the same language: “water” in English does not mean the same as “water” in Twin-English.

\textsuperscript{20} See my discussion of “The Problem of Functional Indeterminacy” below.
having Intentional states that are directed toward the same object(s)—does not tell us what the twins’ Intentional states are *about*. In other words, knowing that the twins’ mental states are directed toward H₂O, XYZ, and whatever else fits their primed definitional script, does not tell us which object among these objects actually activated or primed their scriptD, which *o* lies at the other end of the C relation. From this we may conclude that *directedness does not determine aboutness*.

**Fregean Cases** Fregean cases, on the other hand, demonstrate that knowing what object an Intentional state is about does not tell us what scriptD that object may prime in a given information-processing system. If for Eva there is one real-world object (Venus) that, depending on the time of day, normally primes in her one of two distinct definitional scripts—MORNING-STARₐₛ and EVENING-STARₐₛ—then she might have a belief directed toward the Morning Star without having the same belief directed toward the Evening Star. Notice that we are here speaking of *directedness*, not *aboutness*: Eva’s beliefs will be about the same physical object, but they will be directed differently. The reason, once again, is that Intentional directedness is determined by scripts, and where these scripts differ, directedness differs as well. The fact that Eva’s beliefs refer to the same object is irrelevant; what matters for directedness is the point of view of the information-processing system, not how things actually stand in the external world.

We can assume that Eva’s MORNING-STARₐₛ is distinct from her EVENING-STARₐₛ because some of the frames in one definitional script are incompatible with some of the frames in the other. Her MORNING-STARₐₛ, for instance, might include a frame indicating that it (Phosphorus) only appears in the east at sunrise, while her EVENING-STARₐₛ might contain a frame indicating that it (Hesperus) only appears in the west at sunset. This frame difference means a difference in the directedness extensions of the two scriptDₛ, which ultimately amounts to a kind of *error*: Eva misrepresents the Morning Star as a distinct celestial body from the Evening Star.

Here again is Kripke’s take on the matter:

When the mythical agent first saw Hesperus, he may well have fixed his reference by saying, ‘I shall use “Hesperus” as a name of the heavenly body appearing in yonder position in the sky.’ He then fixed the reference of ‘Hesperus’ by its apparent celestial position. Does it follow that it is part of the meaning of the name that Hesperus has such and such position at the time in question? Surely not: if Hesperus had been hit earlier by a comet, it might have been visible at a different position at that time. In such a counterfactual situation we would say that Hesperus would not have occupied that position, but not that Hesperus would not have been Hesperus. The reason is that ‘Hesperus’ rigidly designates a certain heavenly body and ‘the body in yonder position’ does not—a different body, or no body might have been in that position, but no other body might have been Hesperus (though another body, not Hesperus, might have been called ‘Hesperus’). (1980: 57-8)

In every possible world, everything that is true of Hesperus is also true of Phosphorus. This is what it means for these names to be rigid designators. Why, then, is “Hesperus is Hesperus” not very informative, while “Hesperus is Phosphorus” is? This is Frege’s famous paradox of identity, which is closely related to the failure of substitution of coextensive terms in belief contexts: the point that “a = b” may be informative whereas “a = a” is not—even though “a” and “b” are singular terms that refer to the same object—is in effect the point that one may fail to belief that a = b while believing that a = a. Frege’s well-motivated solution was to posit a sense, different from reference, and to make this sense susceptible to context and manner of description.\(^{21}\)

\(^{21}\) Perhaps it is Kripke’s insistence that a name designates the same object in every possible world—making “The Evening Star is not the Morning Star” a necessary falsehood—that makes it hard for us to envision any situation where Eva might think that the names actually designate different objects. In “Indexical Belief,” Stalnaker argues that there is no tension between the idea of rigid designation and the contemplation of possible worlds in which things are not as they are, just as long as we turn focus from semantical rules to the perspective of the agent. We should ask not just, what does “The
Chapter VII  A Script for Intentionality

The reason that scripts do not rigidly indicate is precisely the reason why “the body in yonder position” does not rigidly designate. Scriptal frames arise from episodic experiences and indirect learning, and what the experiencer is exposed to during these episodes are the non-rigid properties of things, which are nevertheless enough for subsequent recognition. We can think of these scriptal frames as “bits of description” if we like, for they function to detect things that present themselves in a certain way. When objects do not present themselves in that way, the agent fails to recognize them; and when the same thing appears differently at different times (without a clear connection between the instances), the agent is liable to form dissociated definitional scripts corresponding to the different ways the object appears. This is what we have called the aspectual or perspectival dimension of Intentionality.

Oedipus’ case is similar to Eva’s. He wants to marry Jocasta, but he does not want to marry his mother. The problem is to understand how these desires can be different given that Jocasta is his mother. Answer: the information that Jocasta is Oedipus’ mother is not represented anywhere in the information-processing system we call “Oedipus.” In fact, Oedipus has every reason to think of Jocasta as not his mother, since he believes himself to be the natural son of Polybus and Merope (who were, unknown to him, only his surrogate parents). If Oedipus has two distinct scripts, one for MY-MOTHER and one for, say, MY-LOVER, such that the woman whom he wants to marry primes only the latter, it becomes obvious how the mistake took place: despite Jocasta’s maternal relationship to him, she became, through a bizarre sequence of events, causally associated with his MY-LOVER, and Merope (a woman biologically unrelated to him, but who is likewise deceived about their relationship) became causally associated with his MY-MOTHER. Since the two definitional scripts have different directedness extensions, Oedipus can have any number of Intentional states directed toward Jocasta without the same states being directed toward his mother. Alternatively, Oedipus can have Intentional states directed toward Jocasta as his lover (because Jocasta primes his “lover” script) without the same states being directed toward Jocasta as his mother (because Jocasta does not prime his “mother” script).

To protest that Jocasta is Oedipus’ mother, so that Oedipus cannot have an Intentional state focused on Jocasta without having the same state focused on his mother, is again to confuse directedness with aboutness. The former (unlike the latter) is determined only by the nature of information scripts provide; matters of actual fact have no bearing on the question. So in order to know how Oedipus’ affections can be directed toward Jocasta but not toward his mother, we must determine what actual or possible objects are picked out by his MY-MOTHER and by his MY-LOVER. These are not the same, which is exactly where the mistake lies: Oedipus misrepresents Jocasta as not his mother, and misrepresents Merope as his (biological) mother.

Eva and Oedipus show us how the aspectual dimension of Intentionality intermingles with directedness. It matters immensely what information about an object is represented by the frames of its definitional script, since this information determines how the information-processing system whose script it is will regard and interact with the object in question. So because Eva’s MORNING-STAR and EVENING-STAR are mutually incompatible, she does not represent what she calls the “Morning Star” and the “Evening Star” as two aspects of the same body. Rather, we have a situation where at particular times she thinks of Venus under its “Morning Star” aspect, at other times under

Evening Star is not the Morning Star” say (according to the semantical rules)—what is its truth-value at various alternative possible worlds—but also what would it say at various alternative possible worlds, and what would the names rigidly designate if it were true in these worlds. “In general, to understand the content of a person’s belief, ask what the world would be like if the belief were correct. What is the world like, according to the person’s conception of the way the world is?” (Stalnaker 1981: 136).
its "Evening Star" aspect, but at no time does she think of the planet as having the two aspects, "Morning Star" and "Evening Star." Oedipus is guilty of the same kind of mistake. He was only able to think of Jocasta from the perspective of a lover, a sexual partner, a spouse, but not from the perspective of a son. The consequences are relatively benign in the first case, more serious in the second.

What determines the aspectual shape of an Intentional state is just that which determines the state's directedness extension: a script $D$ which represents some (but not all) features or properties of an object. The features or properties of the object that are represented in the frames of the script $D$ define the aspect under which that object is treated by the information-processing system. This aspect will then be blind to other non-represented information. Hence, the possibility of misrepresentation goes hand in hand with Intentionality.

7.2 Misrepresentation

Intentional directedness can be thought of as a kind of narrow content that gives us leverage on the viewpoint of the agent. Because the agent's perspective is heavily influenced by his or her ability to discriminate different facets of the environment, we are likely to find in this ability (or lack thereof) the roots of misrepresentation. We have defined $E_d(S)$ in a manner that establishes the range and limits of what can and cannot be distinguished in the world-according-to-the-agent. Accounting for script-based error may now be accomplished by comparing the script-generated directedness extension of a given Intentional state with the state's aboutness extension.

We will be concerned only with script-based misrepresentation in this section. I am not particularly interested (at least for the purpose of this discussion) in the type of misrepresentation that can be traced back to sources outside an agent: faulty instruments, broken indicator devices, deliberate deceptions, and the like.

i. Conditions for Eliminating Script-Based Error

The first step is to identify the conditions under which no script-based errors occur. Suppose that Delilah encounters an object $o_i$ which primes in her $O_{DS}$ and that $O_{DS}$ is in fact Delilah's definitional script of $o_i$ (or, if you like, her concept of $o_i$). Thus whenever $o_i$ primes $O_{DS}$ in Delilah, she will be in the Intentional state $I_s$ of recognizing (i.e., believe that she perceives) $o_i$. In this case, $E_a(I_s) = \{ x \mid x \text{ is } o_i \}$ since $o_i$ stands in $C$ to Delilah; $E_d(I_s)$ will be the set $o_i/\mathcal{R} = \{ o_j \in \Omega \mid o_i, \mathcal{R} o_j \}$—that is, the set of all actual or possible objects manifesting the properties to which the cluster of neuronal structures constituting $O_{DS}$ has the job of identifying.

Now suppose that Delilah's definitional script of $o_i$ is such that $E_d(I_s)$ contains just one element, $o_i$ itself. Here we have a situation where Delilah's $o_i$, script $D$ is so precise that no actual or possible non-$o_i$ answers to all the frames contained in it. When Delilah has a belief about $o_i$, or a desire for $o_i$, or a fear of $o_i$, her Intentional state will be directed toward exactly the same object that the state is about. Whatever else happens, then, we can rest assured that her entire definitional script of $o_i$ will not cause her to misidentify a token of a non-$o_i$ for a token of an $o_i$. So we might call our first condition for eliminating the possibility of script-based misrepresentation bi-extensional equivalence, which simply states that for any given script-based Intentional state, $E_a(I_s)$ must be equivalent to $E_d(I_s)$. 
But the equivalence of $E_a(I_o)$ and $E_d(I_o)$ is not enough. Putnam's original rendition of the Twin-Earth story places the twins in the year 1750 (Earth calendar). At that time, no one on Earth was aware that water is composed of $H_2O$, and scientists on Twin-Earth had not yet discovered that water is $XYZ$. Let us change the scenario to the present, where Adam is a hydrochemist whose knowledge of the properties and composition of water is as good as anyone's on the planet. Modern-Adam's scripto of water is much more sophisticated and elaborate than his eighteenth-century counterpart's: he knows full well that water is composed of two parts oxygen and one part hydrogen arranged in a particular molecular configuration, and he is familiar with at least a half-dozen tests to verify that fact. So Modern-Adam's $WATER_{DS}$ is so precise that no actual or possible non-$H_2O$ substance answers to all its frames. In short, for Modern-Adam any Intentional state involving "water" satisfies the condition of bi-extensional equivalence.

We now imagine that through the course of an ordinary day Modern-Adam encounters two pitchers, one full of $H_2O$ and the other full of $XYZ$. The salient features of $H_2O$ (colourlessness, odourlessness, tastelessness, viscosity factor, etc.) will be sufficient to prime his scripto of water, so he will correctly believe that the first pitcher contains $H_2O$. Likewise, the salient features of $XYZ$ (which are the same as those of $H_2O$) are sufficient to prime his scripto of water, so he will incorrectly believe that the second pitcher contains $H_2O$. Although bi-extensional equivalence is satisfied in this case, we still have a misrepresentation.

What caused Modern-Adam's mistake? The answer is that only a subset of the frames which constitute his $WATER_{DS}$ (most likely those sensitive to physical appearance) were involved in his identification of the contents of the second pitcher. Other frames—those pertaining to molecular structure, for instance—were not involved. We call this a mispriming error: the incorrect priming of a scripto despite the satisfaction of bi-extensional equivalence. Had Modern-Adam been more vigilant (or sceptical) he could have conducted one of the various tests he knew (represented in his memory in the form of instrumental scripts) to verify the substances' identities. As it were, his knowledge that water is composed of $H_2O$ played no part in the process by which he identified the contents of the pitchers, and therein lies the problem: his misprime occurred because his judgment was not based on all of the frames that constitute his $WATER_{DS}$. This suggests a second condition for preventing script-based misrepresentation, which, together with the first, we can state as follows:

Let $A$ be an information-processing system (agent), $o$, an object, $O_{DS}$ a definitional script in $A$ primed by $o$, and $f_1, f_2, \ldots, f_n$ the frames constituting $O_{DS}$. Let $I_o$ be an Intentional state (in $A$) representing $o$, with aboutness extension $E_a(I_o)$ and directedness extension $E_d(I_o)$. We say that there is no possibility of $A$ scriptally misrepresenting $o_i$ iff,

$$[E_a(I_o) \subseteq E_d(I_o)] \& [E_d(I_o) \subseteq E_a(I_o)]$$

(1) All $f$s constituting $O_{DS}$ contribute to $A$'s identification of $o_i$. Alternatively, the priming of $O_{DS}$ by $o_i$ is a consequence of the combined indicatory results of all constituent $f$s of $O_{DS}$.

22 This just means that $E_d(I_o)$ and $E_a(I_o)$ are equivalent, since two equivalent sets are, by definition, subsets of one another.
ii. Types of Scriptal Misrepresentation

Based on the above, we can identify the following categories of script-based errors. I do not mean the list to be comprehensive or the categories to be mutually exclusive, but I do think that they give us a good idea of how a script theory of Intentionality can handle common mistakes.

**Overgeneralization**

Overgeneralization is one possible consequence of the failure of bi-extensional equivalence. It is what we get when one’s script is “too sketchy,” when $E_a(I_a)$ is a proper subset of $E_d(I_d)$. Note that for any $o_i$ and an Intentional state representing $o_i$, $E_a(I_a)$ is a proper subset of $E_d(I_d)$ whenever an information-processing system cannot discriminate $o_i$ from superficially similar objects. In such cases, even though $o_i$ may be the only object causally associated with $O_{DS}$, $O_{DS}$ will define a directedness extension that includes not only $o_i$, but all other objects the system cannot distinguish from $o_i$. This sort of overgeneralization is exemplified in the original Twin-Earth scenario: the Adam-twins share the same WATER$_{DS}$ which is not fine-grained enough to distinguish $H_2O$ from XYZ. Hence, for both twins any Intentional state involving “water” will be such that condition $(n)$ fails: $E_a(I_a) \neq E_d(I_d)$ because $E_a(I_a) \subset E_d(I_d)$.

**Mispriming**

Modern-Adam’s error is a consequence of the failure of $(n)$, but in one sense it can be viewed as a special case of overgeneralization, since he identified the contents of the two pitchers using only part of his water script. Had all of the frames of his WATER$_{DS}$ been implicated in the identification process, the error could have been avoided. Why? Because Modern-Adam’s script of water is precise enough to exclude all non-$H_2O$ substances from the directedness extension it defines. As it turned out, the part of his script that was used to identify the contents of the two pitchers was not sufficiently fine-grained to distinguish $H_2O$ from XYZ. The result was an overgeneralization error similar to Eighteenth-Century-Adam’s.

But it is important to realize that Eighteenth-Century-Adam had no way of distinguishing between $H_2O$ and XYZ, that $H_2O$ and XYZ were identical under his “water” script. We can say that Eighteenth-Century-Adam was simply a victim of his ignorance, or perhaps the ignorance of the age in which he lived. In any case, correcting ignorance-induced mispriming is never a straightforward matter, since doing so must involve the acquisition of knowledge. In terms of script theory, the correction would require supplementing, or extending, or otherwise refining the definitional script of the object in question. Why does this help? Because the more precise a script is, the more exclusive its directedness extension becomes; and this allows more discriminations to be made.

None of this is true for Modern-Adam, however, whose script does contain the information needed to differentiate the two substances: he knows the molecular structure of water, and his knowledge of various chemical analyses should aid him in making the distinction. So why didn’t he conduct the proper tests? The answer depends on what we might call the entry conditions for a script. Roughly, these are the conditions which are sufficient to prime a given script. I think any explanation should acknowledge the pragmatic utility of Modern-Adam’s extending his belief to include the contents of both pitchers: XYZ certainly looks, feels, tastes, and smells like $H_2O$, so he has no reason to doubt that it is $H_2O$. It would make no sense for him to insist upon conducting a chemical analysis every time he turns on the faucet or takes a shower.

Script theory can account for this sort of pragmatism in the following way. We know that as a hydrochemist Modern-Adam has an unusually detailed WATER$_{DS}$, some of the frames of which pertain to appearance (taste, smell, colour, consistency); some have to do with functional properties (water can be used to make tea, to make ice, to wash a car, to boil a lobster); and some frames will
include specific information about the chemical properties of the liquid (its atomic composition, how it reacts when it comes in contact with sodium hydroxide, etc.). These frames did not simply spring from thin air. They evolved from the numerous episodes experienced by Modern-Adam in which water appeared and had a place. This, after all, is how scripts are constructed: they are products of episodes, experiences, and indirect learning (which also takes place in the context of episodes). Clearly, then, episodic scripts are heavily context-based in that the events contained therein happen to particular people, in particular places, at particular times, in particular circumstances; and that is how they are encoded in memory. So it is not unreasonable to suppose that the frames constituting a given script of that object may be sensitive to the context in which they were formed.

If this is correct, the question of whether or not the features displayed by an object are sufficient to prime a script of that object will depend on the circumstances in which these features are displayed. So, on the one hand, Modern-Adam became acquainted with the appearance and functional properties of water within the course of his day-to-day life. In a normal everyday context, then, substances which display water-like qualities will prime his script; he will note that the stuff that fills his bathtub can be used to do his laundry, just as he will not doubt that the contents of the second pitcher (actually containing XYZ) is H2O. On the other hand, the more technical details were most likely learned in a scientific or experimental environment, and so will become much more important in these settings. Thus we can reasonably guess that in a delicate experiment requiring the precise mixture of a number of compounds, H2O among them, Modern-Adam will not reach for the first colourless, odourless substance he sees. He will first assure himself, using whatever means at his disposal, that the liquid is in fact water. In this situation, the H2O frame of Modern-Adam’s script will be more important than frames sensitive to physical appearance.

Overspecification is another possible consequence of the failure of bi-extensional equivalence. It occurs when the real or actual members of $E_d(I)$—that is, $E_d(I) \cap \mathcal{W}$, where $\mathcal{W}$ is the set of all physical objects—constitute a proper subset of $E_a(I)$. In order for this to happen, there must be a situation where a group $G$ of several distinct objects are connected to a script that defines a directedness extension whose actual members are only some of the objects falling under $G$. We would then have a case where $\eta$ fails because $[E_d(I) \cap \mathcal{W}] \subset E_a(I)$. Here the directedness extension of an Intentional state involving $G$ overspecifies the group of real-world things making up $G$, thus failing to represent some of $G$’s elements as elements of $G$. This sort of error occurs when, for instance, one’s script has a frame like, “all tigers have yellow stripes,” thus failing to represent albino tigers as tigers.

But there is another side to overspecification. Based on my experience with just a few tokens of an object-type $o$, I may form the belief that all tokens of $o$, seen and unseen, have one or another property. The overspecification in this case is a kind of induction, something that is constantly (and often advantageously) practiced in our daily lives. A bad experience with a Rottweiler may “open my eyes” to the violent temperament of this breed of dogs, and a mishap with a kerosene lamp may induce me to be more careful with them in the future. But although overspecification may not always

---

23 Whatever constitutes a physical object in the common way is what “physical object” means here. So chairs, tables, radio antennas, mountains, octopi, aunts and uncles, all go into $\mathcal{W}$. We won’t be sidetracked by technicalities about whether parts and segregates constitute physical objects in their own right. Common-sense will arbitrate problems of this kind: if a part can be separated from a whole, then we can, if we wish, conceive of that part as a unique physical object; if it can’t, then the whole will be the physical object. One thing should be noticed, however. $\mathcal{W}$ is a finite set. Even if we go all the way down to the 0-level of “fundamental” particles and conceive of these as the physical objects in question, $\mathcal{W}$ would still be finite, since the number of such particles in the universe and the number of permutations in which these particles may be combined, though astronomical, is still limited.
be a bad thing, it is nevertheless a good place for error to creep in, which is what we are concerned with here.

Misidentification  Misidentifications are essentially cases of “mistaken identity.” They occur when $E_a(I_a) \cap E_d(I_d) = \emptyset$—that is, when there is no $o$ that is both a member of the $E_a(I_a)$ and a member of the $E_d(I_d)$ of an Intentional state. An example of this type of error is Eva’s mistaking Adam for Samson on a foggy night, in which case Eva’s belief will be directed toward Adam, but about Samson. Oedipus’ confusion is another example. In this situation, Oedipus’ mother (Jocasta) primes in him a script $D$ that misidentifies her as not his mother, and a person who is not his mother (Merope) primes in him a script $D$ that misidentifies her as his mother.

Mirages, Optical Illusions, Hallucinations  These are special cases of misidentification. The error in such circumstances generally involves the “detection” of a particular object that does not exist. Though mirages, optical illusions, and hallucinations give rise to directed Intentional states (e.g., “There is a pond up ahead”), these will be states with no aboutness (there is no actual pond up ahead). Thus the aboutness extension of such states will always be $\emptyset$, and therefore, $E_a(I_a) \cap E_d(I_d)$ will always be $\emptyset$.24

iii. Misrepresenting Hypotheticals

Eva’s belief that there is a colony of Leprechauns living 60 miles northwest of Dublin may also be considered an instance of misidentification, since in this case $E_a(I_a) = \emptyset$, $E_d(I_d) = \{ l \mid l$ is a Leprechaun25$, and so $E_a(I_a) \cap E_d(I_d) = \emptyset$. Eva’s mistake is not that she misrepresents some property that Leprechauns are supposed to have—e.g., their appearance, habits, etc.—but that she misrepresents Leprechauns as real entities when they are not. Under these circumstances, we are perfectly justified in comparing the $E_a(I_a)$ and $E_d(I_d)$ of Eva’s belief in order to reveal her misconception.

But we also want to say that it is possible for someone who knows that Leprechauns are fictitious to misrepresent some of the properties these creatures supposedly possess. For instance, Eva might know that Leprechauns are products of folklore, but erroneously think that they have blue skin.26 How shall we handle this sort of error? We cannot deal with it by comparing the directedness extension of Eva’s belief with its aboutness extension, since there is nothing to which Eva’s belief is about, no real object in the world that we can use to reveal Eva’s mistake.

On the other hand, there is folklore. Eva’s concept of Leprechauns is not original. It is derived from some earlier source that presumably provides a reasonably precise description of these creatures. It is against that source, whatever it might be (and Eva need have no idea what it is), to which Eva’s concept must be compared in order to illuminate any misapprehensions she might have. This may not be an easy task, of course. It will be relatively straightforward if the source of the

24 $\emptyset$ is the result of any set intersected with $\emptyset$.
25 As defined by Eva’s LEPRECHAUNs.
26 Smurfs, I believe, are blue-skinned. We may suppose Eva to have confused that property of Smurfdom for a property of Leprechaunhood!
creation is clear and unambiguous. If I were to say that the “Ghost of Christmas Yet to Come” was a bubbly, cheerful sort of fellow, you may quote Dickens to set me straight:

The Phantom slowly, gravely, silently, approached. When it came, Scrooge bent down upon his knee; for in the very air through which this Spirit moved it seemed to scatter gloom and mystery. (1843: 80)

But things become a bit more complicated when the source is not so definitive. Perhaps Leprechauns have historically been described in competing ways, or perhaps the early history of the myth has somehow been lost. In such cases, social conventions will have to assume the role of the source, thus functioning in exactly the way that Burge envisioned them to function in relation to all propositional attitudes. If it happens that social conventions are equally divided, we may find no systematic way of identifying errors in representing hypotheticals. I don’t look upon this as a particularly unsettling possibility, since nothing so important will hinge on “getting things right” in any case.

I should perhaps point out that the errors discussed above are meant to be the general categories of misrepresentation which script theory predicts. Just as an ICBC researcher can intelligently predict the number and type of motor accidents we may expect next year, even though (s)he is not in a position to speak of any of the details (e.g., speed of vehicles, time of day, state of the drivers, road and weather conditions, etc.) that may figure importantly in the explanation of a given accident-token, script theory can predict the type of error we may expect, even though some of the details that are important to the explanation of particular error-tokens (e.g., the relationship between a script and the properties it represents in the external world) are impossible to anticipate. But the categories themselves are not impossible to anticipate; and it is the nature of these categories on which this section was intended to shed some light.

7.3 Consequences for Externalism

What can we say about anti-individualism in light of the scriptal approach to Intentionality we have constructed? We have already said quite a bit, but there are some loose ends that need to be tidied up.

i. Physical Environment and Linguistic Usage

“Cut the pie anyway you like,” Putnam famously said, “meanings just ain’t in the head” (1975: 144). The meanings of what exactly? Robert Stalnaker tells us, using aluminium and aluminium instead of water and water:

We are invited to note that despite the intrinsic similarities of the two doppelgangers, their utterances have different semantic properties. When the earthling says “Aluminium is used in the construction of airplanes,” she says something that differs in content from what her twin says when she utters the same sounds. Since what is in the heads of the two is the same, while what they mean when they use certain words is different, the meaning of those words must depend on something other than what is in those heads. (1989: 169; my emphasis)

Stalnaker, I believe, is right. The gap Putnam had in mind was between the meaning of words and the narrow content of psychological states, which is why he despaired of ever finding a connection between the two that would make a difference in the one congruent with a difference in the other. That is all we are entitled to conclude from Putnam’s thought-experiment.
Now compare this to the following from Loewer, which (in my view, at least) typifies what is wrong with many analyses of the Twin-Earth story:

B and his twin earth brother B* each develop cognitive structures. B's is sensitive to the presence of water and B*’s is sensitive to the presence of XYZ. The first has a state with the content this is water and the second has a state with the content this is XYZ even though the two states are as physically alike as you please. (1982: 300)

Except for the first sentence, everything said in this passage is either incorrect or misleading:

B and his twin earth brother B* each develop cognitive structures.

So far so good!

B's is sensitive to the presence of water ...

True (strictly speaking), but misleading. Although some cognitive structures in B are sensitive to the presence of water, these structures cannot distinguish between H₂O and XYZ, so they are also sensitive to waterₐ (XYZ). In fact, the structures in question will respond in exactly the same way to every elements in the set H₂O/R = {α ∈ Ω | H₂O ⊂ R α}—i.e., to every substance which shares with H₂O all of the features to which the frames of B’s WATERₐ are attuned. The same consideration shows why the rest of the statement—"... and B*’s is sensitive to the presence of XYZ"—is also misleading even though, strictly speaking, it is true.

The first has a state with the content this is water and the second has a state with the content this is XYZ even though the two states are as physically alike as you please.

This is incorrect. It is false to characterize the content of either twin’s belief as involving exclusively either H₂O or XYZ, just because the word “water” rigidly designates an object on Earth that is different from what the phonetically identical word rigidly designates on Twin-Earth. From the point of view of B and B*, there is no difference between H₂O and XYZ, so the content of their respective beliefs cannot be described as being directed to one but not the other substance.

Twin-Earth situations thus show us that Intentional states are normally not directed toward an exclusive object, but equally toward the individual members of a set of objects which are indistinguishable under a particular scriptD. This is what it means for Eₐ(Iₐ) to be defined in terms of an equivalence class. It means that an agent standing in a directedness relation with one member of Eₐ(Iₐ) will be standing—without distinction or prejudice—in the same directedness relation with all members of Eₐ(Iₐ). Of course, in reality these members may be as different from one another as H₂O and XYZ, but not from the point of view of the agent, and the point of view of the agent is what Intentional directedness is supposed to capture. Generally, the more precise and elaborate a definitional script is, the smaller will be its directedness extension. An extreme case is one in which a scriptD of some object is so precise that the directedness extension to which it gives rise includes only that single object as its member (as in the example of Modern-Adam).

This situation may be more common than one supposes. How does Samson know that what he sees is a human being when the person in front of him is a complete stranger? The answer, I think, has to do with the fact that Samson’s generic HUMANₐ is broad enough to accommodate (i.e., be applicable to) indefinitely many distinct people, while, at the same time, precise enough to separate humans from non-humans. So even though the woman Samson meets is someone whom he has never seen before, she will nevertheless manifest features with which Samson is very familiar: she is bipedal; she has two eyes and two ears; she is attired and speaks some language. These and other
properties are enough to prime Samson’s HUMAN$_{ds}$, which will have a lot of “empty slots” waiting, as it were, to be filled by her unique characteristics.

It will not take Samson long to fill these slots: he will note that the woman in front of him is about 5’6”, 125 pounds, with blue eyes and blond hair, a fair complexion, and is named Delilah. Barring the kind of scepticism that legitimizes the possibility of radical sensory deception, it is hard to imagine what object, other than Delilah herself, Samson’s newly formed DELILAH$_{ds}$ may be directed toward during the time of their initial meeting. Of course, later on that day, when Samson reflects back on the encounter and remembers Delilah as having this or that property, the directedness of his Intentional state will reflect his discriminatory limits. Does this mean that we can never think of particular objects? Of course not, since Intentional states have aboutness just as they have directedness. But if it is only directedness we are concerned with, then we have no reason to expect our thoughts to show any more precision than what our cognitive systems allow. Alas, we are not omniscient after all.

I hate to be the bearer of bad news, but so it must be. Some have found this unsatisfactory and have sought refuge in an externalist conception of memory. Thus Peter Ludlow claims that “it seems natural for the externalist to reason that if the contents of our mental states are determined by external conditions, then the contents of our memories are subject to the same external conditions. Accordingly, it is possible to construct twin cases in which two psychologically identical agents have different memories (say one has memories of drinking water, and one has memories of drinking twater). Likewise, if ... the contents of our mental states are determined by our social environment, it is natural to suppose that the contents of our memories will depend on our social environment ... The consistent social externalist is bound to say that the content of a memory is fixed at the time recollection takes place ... This view of memory ... implies that the contents of our memories will shift over time as the fabric of our linguistic community changes” (1995b: 308-9). Frankly, I am not sure why anyone would accept such a position as somehow more palatable than one which limits the precision of Intentional directedness to the range of discriminations an information-processing system is able to make. At any rate, the onus is on the externalist. If we are to conceive of memory in a manner that is so repugnant to the way in which psychology has understood it for so long, we must at least be given good reasons, reasons which have not yet been made explicit.

On the other hand, there appear to be cases where the $E(I)$ of some Intentional states are naturally congruent with their $E(I)_s$. In an article called “Intentionality and the Non-Psychological,” C. B. Martin and Karl Pfeifer argue that “if someone wants his rubber ball, he wants that rubber ball and not just anything exactly like it” (1986: 551). I agree. It is possible for Adam to have a desire expressible by the form, “I want the rubber ball I was playing with yesterday and no other,” even though today he cannot distinguish between the object in question and a molecule-for-molecule replica of it. The precision of Adam’s Intentional state is here explained by the manner in which the object of his interest is identified: it is picked out entirely in terms of the causal relation $C$ which connects Adam uniquely to a particular object. Thus the way in which Adam can genuinely desire his ball, despite the fact that he cannot distinguish it from another exactly like it, is by deliberately letting $C$ do most of the identificatory work on his behalf. This, of course, must involve a suspension of judgement on Adam’s part: he must admit that he cannot make the required distinction, and he must find a way to specify what he wants in a manner that does not depend on his ability to differentiate the possible candidates.

The above shows that our scriptal approach is by no means radically internalist. It does, however, claim a little more than what a purely externalist position permits. Tyler Burge (1979) has argued that it is not just meaning and other semantical properties, but also Intentional psychological
properties that are shown to depend on external conditions. In particular, facts about the linguistic
practices of members of the agent’s community are among the external conditions on which
Intentional mental states depend. Thus Adam’s doppelgänger on Twin-Earth not only means
something different from what Adam means when he uses the word “water,” Adam also expresses a
different belief (a belief with a different content) when he says “there is water in the pitcher.” The
lesson is not that there is a gap between speakers’ linguistic practices and their mental states, but that
mental states themselves have to be understood partly in terms of the way a person interacts with his
or her environment.

In “Social Content and Psychological Content,” Brian Loar stands firm against this sort of
reasoning. He holds on to what I have suggested is the moral of Putnam’s Twin-Earth scenario—that
there is a difference between the meaning of words and the content of Intentional states. “I shall
argue,” says Loar, “that psychological content is not in general identical with what is captured by
oblique that-clauses, that commonsense constraints on individuation induce only a loose fit between
contents and that-clauses” (1988a: 102).

The problem with externalism as I see it is this: if the claim is just that mental states have to
be understood partly in terms of the way a person interacts with his or her environment, there would
not be much of a controversy. Most people acknowledge as much, and the scriptal approach I
recommend certainly respects the role of the external world: it not only posits an aboutness
dimension to Intentionality, but also admits evolution and individual development as forces that help
to form a person’s narrow psychology. The problem is in overemphasizing environmental relations,
or worse, making them exhaustive of the kind of thing Intentionality is supposed to be. Is it really
difficult to believe that Adam might actually be involved in shaping his own mental life? Is it hard to
imagine that he might infect the character of his Intentional states with his perspective or point of
view? If perspective counts, if it makes any difference at all to the nature of thought, then we really
have no choice but to find a way of making sense of it. This is what script theory does so neatly.

It is thus open to the internalist to acknowledge a wide or broad content—one that
individuates mental states in terms of the thinker’s social and physical world—as well as a narrow
content that calls attention to the way things seem from a first-person standpoint. Burge expounds his
externalist position using a number of very clever examples, all of which make use of the same
argument-form: one conceives of a person in a certain mental state or experiencing a certain mental
event. Then holding that person’s individualistically, non-Intentionally specified states constant, one
conceives of a different environment that presents the person with the same proximal stimulations as
the first environment but in which the person allegedly lacks the mental state or experience that he
had in the first environment. One concludes that individualism is false, inasmuch as the thought-
experiment is taken to show that the individuation of a person’s mental states are a function not
simply of the person’s individualistically, non-Intentionally specified states, but also of the
environment in which the person is embedded.

But where is the perspective of the agent in all of this?

To my mind, Burge never really succeeds in motivating a reason for why we should extend
the reach of environmental influence to every aspect of Intentionality. Where the environment really
does have a say is in those areas which are (pardon the expression) “public domains,” natural

---

27 See Bilgrami’s (1988) comments on Loar’s article, and Loar’s (1988a) reply to Bilgrami.

28 Discussed in §3.2.ii.
language being an obvious example. LASCH gives us one way to think of the connection between language and thought, but the hypothesis is also consistent with the setting of some limits: as long as we can think without vocalization, as long as our cognitive liabilities limit the way we apprehend things, there will always be subjective points of view.

Some have dismissed this kind of sentiment as resting on “a false conception of mind—the Cartesian conception that continues to bewitch us despite the widespread criticism to which it has been subjected” (Stalnaker 1993: 213). But while I am sure it must be comforting to shroud all those sympathetic to the idea of narrow content with a Cartesian cloak, I think this practice should be avoided. The heyday of substance dualism (if there ever was one) is long past, and it is perfectly possible to construct a naturalized notion of narrow content that is equal to longstanding intuitions without being committed to Cartesian doctrine. There is nothing mysterious about narrow content. It is simply a name we have given to an obvious fact: the way people perceive things is constrained in specific ways by their “instruments of representation.” That is what it really comes to. And what is the significance of this fact? Only that we had better get some grip on what these constraints are if we ever want to understand how people see the world they represent.

ii. Causal-Evolutionary History

Stalnaker’s Externalism The Adam Twins have often been taken to show that the causal or evolutionary history of an agent contributes essentially to determining his or her Intentional content. The reason Adam’s thoughts are about H₂O but Adamₑ’s are about XYZ, it is argued, is due to corresponding differences in the twins’ biographies: Adam’s history connects him to H₂O; Adamₑ’s connects him to XYZ. Such a view is advanced by Stalnaker, who seems to understand Putnam’s argument in “The Meaning of ‘Meaning’,,” but nevertheless shares Burge’s impression that the moral of Twin-Earth can be applied to the nature of thought directly.

Stalnaker’s basic idea about what Intentional states are is closely related to Dretske’s:

The theoretical account I have in mind is the information-theoretic account of Intentional content. The rough idea is this: states of mind carry information when there exists a pattern of counterfactual dependencies between those states and corresponding states of the environment. If x is in a state caused by the fact that P, and would not have been in that state if it had not been that P, then that state of x carries the information that P...

According to this account, some of Adam’s internal states are beliefs about water because they are normally sensitive to facts about water. What makes it true that Adam’s internal states depend on facts about water is that water has certain observable properties, and is normally the only stuff around him that has those properties. These facts do not obtain on Twin-Earth, and that is why parallel considerations show that Adamₑ’s internal states, though physiologically identical to Adam’s, carry information about XYZ, not H₂O.

Notional Attitude Psychology Once things are described in this way, it may become difficult to see why narrow content is necessary at all: we can, it seems, define a mental state
externally and account for misrepresentation by using a context-dependent notion of "normal" that
brings in facts about the causal-evolutionary history of the organism. Stalnaker (1989) thus attacks
Dennett’s (1982) attempt to isolate the “organismic contribution” to the content of its beliefs. Dennett
calls his idea notional attitude psychology, and contrasts it both with propositional attitude
psychology (which describes attitudes in terms of the ordinary wide conception of content) and
sentential attitude psychology (which takes the contents of attitudes to be sentences of an inner
language).

The task of notional attitude psychology, according to Dennett, is to explain how the purely
internal properties of an organism can be used to pick out a set of possible worlds compatible with
the organism’s Intentional states. For example, Adam believes correctly that there is water in the
pitcher. The proposition he believes is true on Earth (where “water” rigidly designates H₂O) but false
on Twin-Earth (where “water” rigidly designates XYZ). Dennett, however, suggests that there is a
different proposition which does not distinguish the actual world from the counterfactual world—a
proposition that we might roughly describe as the proposition that there is some water-like stuff in
the pitcher. The first proposition is the wide content of Adam’s belief. The aim of Dennett’s project
is to define a narrow content so that the second of these propositions is the narrow content.

Isolating narrow content using Dennett’s method does not require reference to causal history.
Suppose we know everything about an organism’s capacities and dispositions, but nothing about
how it got this way, nothing about its historical properties or about the environment in which it grew.
How do we go from this limited information to a characterization of the organism’s notional content?
“Our task,” says Dennett, “is like the problem posed when we are shown some alien or antique
gadget and asked: what is it for?” (1982: 155). We cannot know what it was actually designed for, but
we can try to ascertain, from its internal properties, what functions it is ideally suited to perform.
“We try to imagine a setting in which ... it would excellently perform some imaginable useful
function” (1982: 155). In the same way, to find Adam’s notional world—or the world-according-to-
Adam—we try to imagine the environment (or class of environments) for which Adam as currently
constituted is best fitted. Propositions true in those environments will be the narrow contents of
Adam’s beliefs.

Stalnaker, in contrast, is committed to an information-theoretic account of Intentionality that
relies on patterns of counterfactual dependencies between Intentional states and corresponding states
of the environment: “If x is a state caused by the fact that P, and would not have been in that state if
it had not been that P, then that state of x carries the information that P.” The trouble is that the
Twin-Earth story makes this condition seem too strong: Adam’s belief that water is in the pitcher is a
state that is caused by P (water), but it is also true that he would have had this same belief (only
falsely, in this case) had it been waterₑ in front of him. So Stalnaker’s counterfactual dependency
criterion fails; and since we can put virtually anything in place of water in the Twin-Earth story, it
seems that no x can ever carry the information that P.

Stalnaker’s answer is to invoke a notion of normalcy—“a state represents the world as being
such that P, and so is a state with informational content P, if and only if under normal conditions it
would carry the information that P” (1993: 214)—and he grounds this notion in the causal-
evolutionary history of the organism. Thus it is hardly surprising that Stalnaker admonishes
Dennett’s notional attitude psychology as something that “doesn’t look like what we want at all.
Possible worlds picked out in this way look more like worlds in which the organism’s needs or wants
are satisfied than like worlds in which its beliefs are true” (1989: 182). Stalnaker insists that beliefs
are states that help the believer to cope with an environment, and that the contents of these states are
essentially connected with the kind of environment they help the believer to cope with.
“So be it,” one is tempted to say, “Wherein lies the problem?” What is it about trying to understand the world-according-to-Adam that abandons the idea that Adam is connected to his environment? The idea that Adam’s intentional states must help him navigate the world does not entail that the world-according-to-Adam must therefore be perfectly congruent with the world-as-it-is; what it does suggest is that the incongruence cannot be evolutionarily relevant. So the fact that Adam does not distinguish \( \text{H}_2\text{O} \) from \( \text{XYZ} \) tells us something very important about the kind of information-processing system he is, but it does not undermine his ability to be guided by his “water” beliefs, because there is no \( \text{XYZ} \) on Earth.

The Problem of Functional Indeterminacy Despite Stalnaker’s criticisms, I think Dennett is onto something important. He points to one of two serious challenges facing causal-evolutionary theories of content. Call the first the problem of functional indeterminacy.\(^{29}\) The following from Millikan puts the problem in the right perspective:

If I can make it plausible that the entities that folk psychology postulates are indeed defined by their proper functions, and make plausible that the proper functions with which folk psychology endows these entities very likely are had by some special parts or states of the body, that should be enough to show that cognitive science can probably use folk psychology as a starting point. The job of cognitive science would then be, in part, to explain what the Normal constitution of these psychological entities is and how they Normally perform their defining proper functions. (1986: 61)

Millikan believes that to understand Intentionality, we must look upon the brain as a tool that performs a proper function in “Normal” conditions, where “propriety” and “Normalcy” are again grounded in the evolutionary development of an organism.\(^{30}\) But now concerns about functional indeterminacy become rather pressing. Let me illustrate with an analogy.

Suppose a mechanically inclined gift-shop owner decides to build a gift-wrapping machine in an attempt to streamline his business and improve customer service. He does so, and finds the machine very useful: after being loaded with wrapping paper and boxes of various shapes and sizes, the machine would set about wrapping the boxes quickly and efficiently. Despite the obvious differences, we can think of the shopkeeper as having assumed the role of evolution: just as species evolve because of certain selectional pressures in their environment, thereby acquiring “biologically proper” functions, we can think of the machine as having been designed in response to certain pressures in the store’s environment. We can then say that the machine has been given a “mechanically proper” function—the gift-wrapping job for which it was designed.

But now consider: the machine can perform other tasks, and with no more resources than those available to it by virtue of its original design. For instance, the gift-wrapper can be used to prepare small parcels for the post office by wrapping them in brown paper. It can also be used to prepare the mail-order catalogues for the store’s off-site clientele. In light of these “newly discovered” abilities, can we still consider the wrapping machine to have a mechanically proper function?

\(^{29}\) This is something that Dretske (1986) has talked about at some length. I’ll get to Dretske shortly.

\(^{30}\) Millikan distinguishes what is Normal (note the capital “N”) from what is statistically normal (note the small “n”) or usual. A device operates Normally in a Normal environment only when it is doing its job properly, and this need not occur very often. A sperm cell, for instance, is statistically unlikely to find itself in the Normal circumstances for the fertilization of the ovum. Nevertheless, the fertilization of ova is what sperm is for.
I think not. Part of the reason is that the machine, *as designed*, cannot be a gift wrapper without being also a parcel wrapper. So all that remains to identify the "proper" function of the machine is the original intention of the designer. Millikan (1984) thinks this is enough. She believes that entirely new things in the world, things with no evolutionary history, can have functions *derived* from their "producers," if the new thing was created in order to perform a specific job. But I want to suggest that the shopkeeper's intention has less of a claim to being the determiner of the function of the machine than the machine's physical constitution. Thus if someone were to insist that the machine's proper function is to wrap *gifts* because that is what the shopkeeper *had in mind* when he designed it, one cannot help but feel a certain arbitrariness in this judgement. The whole idea of a mechanically proper function seems, in this case, to be an artificially imposed classification that does not properly describe *what the machine can actually do*. The gift-wraper might as well be a parcel-wraper, for it can do both jobs equally well. So while it is quite obvious that the machine has a design, and obvious that it has various capabilities stemming from this design, it is not so clear that the machine can be said to have a mechanically proper function, at least not if such function is to be fixed by the shopkeeper's will.

If you want to know what the machine can do, you must somehow extract it from the history of its development. Finding out what the machine's physical constitution allows it to do reveals its functional properties without the kind of arbitrariness or prejudice which the shopkeeper's purpose brings into the matter. We would then discover that the *state of wrapping a gift* is functionally identical to the *state of wrapping a parcel*, so that if the machine had a "point of view," it would not be able to differentiate these two states. It is doing just as much what it is supposed to do when it is wrapping gifts as when it is wrapping parcels. That is not to say that the shopkeeper's original desires or goals are irrelevant absolutely, but it does suggest that they are irrelevant to what the machine *once designed* is able to accomplish.

The same is true for human cognitive mechanisms. In fact, in the case of humans the point becomes more urgent, since we cannot even speak of the process of natural selection as having anything like the intentions of the shopkeeper. We can agree that people have the biological design they have because of certain selectional pressures in their evolutionary history, and that due to these pressures the human brain has acquired the capacity to map the world accurately. But the project of finding out what certain human brain structures do—insofar as such functions bear on the perspective of agents—is neither compromised by normative considerations nor restricted to causal-evolutionary facts. Dennett tells us exactly how to extract the functional properties of the brain from its past: consider the brain as a novel artefact and find out what it can do from the way it is put together. Now you know something very profound about the organism whose brain it is; now you know how this organism represents part of its world.

Suppose now we ask, what exactly is Adam's WATERD5 supposed to do? The causal-evolutionary reply is a familiar one: Adam's WATERD5 picks out water (H₂O) because that is what the scriptD0 was designed to do, and because Adam's history ties him to that particular liquid and not any other. But why can't we say that the function of Adam's WATERD5 is to pick out exactly those substances which manifest all the features to which the frames of his "water" definitional script are sensitive? After all, Adam's WATERD5 picks out H₂O by *picking out a certain set of salient characteristics of the liquid* and, therefore, picking out those characteristics is as much a fact of Adam's evolutionary development (and local learning episodes) as the selective pressures his ancestors were under. So it would seem that Adam's WATERD5 is doing *exactly what it is supposed to*

---

31 Dretske (1986) makes a similar point.
do when it reacts to any member of the set $H_2O \Omega = \{ o_j \in \Omega \mid H_2O \mathcal{R} o_j \}$, including XYZ. Moreover, the fact that Adam’s WATER reacts indiscriminately to the members of this set, but is still causally connected only to $H_2O$, can also be explained by citing evolution. Simply put, the distinctions in question did not have to be made, because no environmental pressures required Adam’s script of water to perform its indicatory function rigidly: there was no XYZ on Earth (or any other water-like substance) whose similarity to $H_2O$ might have hampered human evolutionary progress or confused Adam and his progeny into thinking that it is water they see (in rivers, lakes, rainfall, and such) when it wasn’t.

Dretske considers just such an argument in “Misrepresentation,” but does not, I think, give a very strong reply to it.

Suppose an organism ... has two ways of detecting the presence of some toxic substance $F$. This may be because the organism is equipped with two sense modalities each (in their different way) sensitive to $F$ (or some modally specific natural sign of $F$), or because a single sense modality exploits different external signs (or symptoms) of $F$. As an example of the latter, consider the way we might identify oak trees visually by either one of two ways: by the distinctive leaf pattern (in the summer) or by the characteristic texture and pattern of the bark (in winter). We have, then, two internal states or conditions, with an $h_1$ and $h_2$, each produced by a different chain of antecedent events, that are natural signs of the presence of $F$. Each means, that $F$ is present. Suppose, furthermore, that, having a need to escape from the toxic $F$, these internal states are harnessed to a third state, call it $R$, which triggers or releases a pattern of avoidance behaviour. Figure 1 assembles the relevant facts. $R$, of course, is also a natural sign of $F$. Under normal circumstances, $R$ does not occur unless $F$ is present. $f_1$ and $f_2$ are properties typical of normal $F$s. $s_1$ and $s_2$ are proximal stimuli.

If, now, we present the system with some ersatz $F$ ... something exhibiting some of the properties of real $F$ (say $f_1$), we trigger a chain of events ($s_1$, $h_1$, $R$ and avoidance) that normally occurs, and is really only appropriate, in the presence of $F$. If we look at the internal state $R$ and ask what it means under these deceptive conditions, we find our selves unable to say ... that it means anything short of (i.e., more proximal than) $F$ itself. Even though $s_1$ (by means of $h_1$) is triggering the occurrence of $R$, $R$ does not mean (hence, cannot mean) that $s_1$ (or $f_1$) is occurring. $R$ is analogous to a light bulb connected to switches wired in parallel either of whose closure will turn the light on. When the bulb lights up, it does not mean that switch no. 1 is closed even when it is this switch’s closure that causes the light to go on. It does not mean this, because there is no regular correlation between the bulb lighting up and switch no. 1 being closed (50 per cent of the time it is switch no. 2).

If we think of the detection system described above as having the function of enabling the organism to detect $F$, then the multiplicity of ways of detecting $F$ has the consequence that certain internal states (for example, $R$) can...
indicate (hence meant) that $F$ is present without indicating anything about the intermediate conditions (i.e. $f_i$ or $s_i$) that "tell" it that $F$ is present. (1986: 139-40)

We have talked about the multiplicity of ways of detecting $F$ before. It was suggested, during our discussion of the autoassociation model of scriptal priming in §6.3, that the activation of a script requires only that the stimulation of its frames exceed a particular threshold. This leaves indeterminate the specific frames involved, thereby permitting many potential ways in which a single script$_D$ may be triggered.

But does this really mean that Adam’s WATER$_{DS}$ must therefore be directed toward H$_2$O exclusively? What Dretske’s argument shows is that Adam’s script$_D$ does not have the function of picking out objects manifesting only a proper subset of the features to which the frames of the script$_D$ are sensitive. But the argument does not show that Adam’s WATER$_{DS}$ lacks the function of picking out objects manifesting all of the characteristics in question, otherwise WATER$_{DS}$ could not be said to have the function of picking out H$_2$O either. The problem is not related to the indeterminacy of the manner in which WATER$_{DS}$ becomes active; it is related to the indeterminacy of the kind of object the script$_D$ reacts to when functioning properly. Dretske’s argument is not helpful because we are not claiming that Adam’s definitional script of water has the job of selecting some set of features or properties, but the job of selecting objects with those features or properties. It so happens that the way Adam’s script$_D$ is put together makes it impossible for it to pick out water without also picking out water$_e$. So the original problem remains: if Adam’s WATER$_{DS}$ cannot distinguish between H$_2$O and XYZ, it is not a simple matter to say which of these liquids it has the function of indicating. Appealing to evolution or causal history to settle the matter only begs the (functional indeterminacy) question at issue.

On the other hand, we might accept Dretske’s analysis as applicable to Intentional aboutness, which posits an external causal relation between an object and a definitional script. Referring to causal history and evolution in this context would explain why a given object is causally associated with a given script$_D$. But this does not take account of the directedness dimension of Intentionality, nor of the world-according-to-Adam in which there is no difference between water and water$_e$. So a purely evolutionary account is insufficient.

**Back to the Swamp** The scriptal approach suggests a view of directedness as dependent on nomic relations between properties of objects and properties of scripts. Adam’s belief is directed toward water because the property of being water is nomically connected to the frames of his water script$_D$ in virtue of what these frames indicate (which, in turn, is a function of their neural architecture). Moreover, for any object $o$, that answers to all the frames of Adam’s script$_D$, the property of being $o$, is nomically connected to Adam’s script$_D$ of water, since his water script$_D$ does not distinguish $o$ from H$_2$O. This means that a nomic relation between an object and a definitional script can be in place even if that object has never been the cause of the script$_D$’s priming.

The same point can be made by invoking dispositions. A given substance might be just as soluble in XYZ as it is in H$_2$O, even if it never encounters XYZ. Analogously, Adam’s belief might be equally directed toward XYZ and H$_2$O, even though he has never been in causal contact with the former. The upshot is that there is just no way that Adam can have a belief directed toward H$_2$O, given the indicatory constraints of his WATER$_{DS}$s, but not toward XYZ or vice versa, not only because he does not know the difference between the two substances, but because he has no clue that there is a difference between the two substances; because, as far as he as an information-processing system is concerned, the two substances are identical.
The cost of emphasizing aboutness, analyzed in terms of causal history, over directedness, analyzed in terms of dispositional nomic relations, brings us to the second major challenge facing causal accounts: Davidson's unfortunate Swampman.

Lightening strikes a tree in a swamp; Davidson is disintegrated, while entirely by coincidence a nearby tree is turned into his physical replica. Swampman behaves exactly as Davidson does so that "no one can tell the difference." But according to the causal-evolutionary view of Intentionality, the difference is that Swampman has no Intentional states at all. No beliefs, no desires, no hopes or fears. This is because he lacks, by assumption, the causal connections to the world (or indeed to any world) on which Intentional content depends.

In The Elm and the Expert, Fodor claims that he heard Ruth Garrett Millikan explicitly endorse this appraisal of Swampman (1994: 126). Davidson agrees and so, it seems, does Dretske. But unlike Davidson and Millikan, Dretske accepts the verdict reluctantly. He does not argue for it, but tries to convince us that a zombie-like Swampman is possible (1995: 148). And so it is. We can acknowledge this without conceding very much. What is required to make the causal view convincing is not only the possibility of a contentless-headed Swampman, but a non-question-begging argument telling us why we should take that possibility seriously.

In fact, we have no good reason to believe that Swampman is in any way different from Davidson. Let us suppose that Davidson is not disintegrated by the lightening bolt, but is rendered unconscious for a short duration of time. When he comes to, he sees another Davidson—a particle-for-particle replica of himself created from the molecules of the stricken tree—also struggling to get to his feet. Each of the two Davidsons is shocked to see the other and does not quite know what to make of the situation, but they agree on some fundamental things: each remembers a lightening bolt hit a tree just before he was rendered unconscious; each remembers seeing an "impostor" staring back at him when he woke up; each claims of himself to be the "real" Donald Davidson, and, once the truth about the tree came out, each (holding a Davidsonian causal account of Intentionality) insists that the other is a zombie who only appears to have genuine thoughts.

Davidson's family does not know how to tell the two apart, but they are naturally keen on finding out which is which. Accordingly, they put together the finest scientific team they can manage—psychologists, sociologists, biologists, doctors and surgeons—together with some "pertinent" laypersons—friends, acquaintances, former students, colleagues, kin. Everyone agrees that the best way to discover the "real" Davidson is to assume that the causal-evolutionary account of Intentionality is correct and then try to figure out which of the two is the zombie.

So they set about with their tests: biologists take blood and tissue samples from the two Davidsons to compare their DNA; psychologists conduct lengthy interviews to try to find any difference between them or to uncover any anomaly that might throw suspicion on one or the other; sociologists do their best to try to chart each of their relationships with their peers and with the organizations to which they belong; doctors and surgeons probe their bodies and compare the results with their records: Do they have this or that scar from a past surgery? Are their dental records the same? Do they suffer from the same ailments? Are they in need of the same medication(s)? Finally, their friends and family are allowed to question them: Do they remember such-and-such an event? What happened that day when ... ? If we accept the claim that "no one can tell the difference," the results of all these tests are guaranteed to be inconclusive.

Shall we now ask what possible reason do we have for accepting the claim that one of the Davidsons does not really have Intentional states? We have conducted all the tests we can think of
and found absolutely no difference between the two. This negative result is not neutral with respect to causal-evolutionary accounts, but serves to undermine them. Do we not have good reason, based on empirical evidence, to reject a necessary link between the biographical life of an organism and that organism’s thoughts and attitudes?

We can look upon the procedures used to test Davidson and Swampman as a very elaborate kind of Turing Test. What makes this Turing Test elaborate is that we are not attempting to determine the identity of the “impostor” only by analyzing the way questions are answered, but by analyzing everything we can think to analyze: chemistry and biology, psychology, personal habits, preferences, social relations, philosophical views. The trouble is that by premise, nothing we could ever do will show that the evidence that Swampman (whoever of the two he might be) thinks and desires, sees and hears, hopes, itches and lusts, is somehow different in kind or inferior in quality from the evidence that Davidson does all these things. In the face of this kind of futility, what sense can we even make of the claim that there still is a difference? And who exactly has the burden of proof?

The problem has forced Fodor, even in his externalistic mood, to admit that Swampman is a “serious embarrassment” for historical accounts of content:

Of course, not having had one, Swampman doesn’t remember his twelfth birthday party; ‘remember’ is factive, and you can’t remember what didn’t happen. But it seems very odd to say that Swampman doesn’t know what time or day of the week it is … If it’s not his believing that it’s Wednesday that explains why the Swampman says “It’s Wednesday” when you ask him, what on earth does? ... To put the point another way: Perhaps it’s true, as it were, by definition that beliefs, desires, lusts and the like are constituted by their histories; in which case, of course, Swampman doesn’t have them. But, so what? It’s intuitively plausible that he has states that are their exact ahistorical counterparts and that these states are Intentional. (1994:117)

The most that Davidson’s thought-experiment can show is that Swampman’s Intentional states have no aboutness, since they are not causally tied to the world in the requisite way. But this does not mean that those states are undirected. Swampman still has a perspective, a manner in which he represents the world to himself, and he has this perspective in virtue of exactly the same neural structures that are operative in Davidson. Since Intentionality does not require aboutness, the Intentional integrity of Swampman’s mental states remains unscathed despite his lack of evolutionary heritage.

7.4 Radical Individualism

The brand of externalism endorsed by Davidson, Dretske, and Millikan derives from their tendency to regard evolutionary history as essential for content. I think this is wrong. But there is nothing in what I have said about Intentionality that is inconsistent with the idea that our history and environmental conditions exert great influence on our mental lives. In fact, the entire notion of scripts is based on a presumption of a complex network of interactions between agent and world. So far, externalist intuitions are well accommodated.

On the other hand, it is important to understand that the environment exerts its influence by proxy, through the representational medium of the physical brain. What gives Intentionality its directedness and aspectual shape are networks of neural structures that are sensitive to specific types of stimuli. What gives these structures Intentional content is that they are thus sensitive, not how they became thus sensitive. Perhaps that is what Fodor had in mind when he suggested that
Swampman could have ahistorical Intentional states. At any rate, Swampman has the same neuronal structures as Davidson, and insofar as these structures constitute scripts in one individual, they constitute scripts in the other. How these scripts came to be is not nearly as important as the fact that they are there.

But because we are products of a long and complex evolutionary history, we should resist radical internalism. To really appreciate the Intentionality of thought, we must come to terms with both its aboutness and its directedness. To that end, we can think of an Intentional state \( I \) in terms of the ordered pair \( \langle A, D \rangle \), where \( A = E_a(I) \) and \( D = E_d(I) \). Doing so permits us to concentrate on one or another Intentional dimension as best suits our investigative aims: we can consider \( A \) independently of \( D \), and \( D \) independently of \( A \). But if our goal is to achieve an understanding of Intentionality pure and simple, \( \langle A, D \rangle \) must be contemplated as a single unit, and in so doing we shall get as close to the exact character of \( I \) as we can ever hope to get.
Philosophy does not often produce stable, reliable "results" the way science does at its best. It can, however, produce new ways of looking at things, ways of thinking about things, ways of framing the questions, ways of seeing what is important and why.


Questions arise from a point of view—from something that helps to structure what is problematical, what is worth asking, and what constitutes an answer (or progress). It is not that the view determines reality, only what we accept from reality and how we structure it. I am realist enough to believe that in the long run reality gets its own chance to accept or reject our various views.

A. Newell, *Artificial Intelligence and the Concept of Mind*, p. 43.

8.1 Where Have We Been?

I did not care much for philosophy, at least not in the beginning. I took my first philosophy course—called "Thinking, Saying, and Doing"—in order to fulfill an undergraduate breadth requirement and to appease a faint sense of curiosity. The course was taught by an eccentric professor whose odd mannerisms and demeanor made philosophy seem alive and dynamic—especially issues having to do with thought, which captured my imagination more than any others. I read the ancients and the moderns, and initially found it hard to navigate the myriad arguments and opinions. I was persuaded by every point made, until someone proposed to contradict it. What eventually emerged from this chaos of indecision were two important discoveries: I found out that I was a materialist with respect to the workings of the universe, and a realist with respect to mental phenomena.

There is nothing particularly unusual about either of these attitudes, of course, but they did produce in me some uneasiness. Though I never looked upon the prospect of a forlorn Dualism as regrettable, I thought it important to try to incorporate, as much as possible, workaday intuitions into the activity of theory construction. To my mind, folk psychology—in virtue of its power, elegance, and simplicity—was the place to start. Thus sprang my central concern: how does one construct a materialist theory of mind that respects folk psychological categories? I was determined to find an answer, yet still prepared for disappointment: should folk psychology turn out to be a sham, we shall have to do our best to swallow that bitter pill.

Things did not get easier, even when I came to realize that many others shared my ambition. The more I thought about it, the more I became impressed with just how difficult it is to transform folk psychology into a more robust doctrine. The problem has two distinct components. First, there is the technical side of actually constructing a mesh of reasonably well-formulated ideas capable of
doing the formal work required, and second, there is the public relations side of having to fend off the usual alarmist charges of “falling prey to Cartesianism” that such an approach would inevitably incite. The latter of these two tasks clearly depends on the former, making it far more important.

My objective in this project was simple. I wanted to investigate what makes some mental states Intentional as well as preserve some strong beliefs I had about the nature of cognition. By the time I started researching my dissertation, the most pressing of these beliefs had become an urgent directive: find a legitimate place for the agent’s perspective in his or her mental life. It seemed to me that fulfilling this directive will have to involve an attempt to construct a naturalistic notion of narrow content that simultaneously legitimizes and demystifies it.

Fortunately, a number of very significant steps had been taken already. I rely on a lot of work done by others in my dissertation. This is something which I acknowledged in the last paragraph of chapter I, citing specifically the labours of Schank-Abelson and Dretske. What I take to be the innovations of my thesis are:

(a) the manner in which I blend the Schank-Abelson notion of scripts with Dretske’s description of simple concept formation; (b) my division of the aboutness and directedness dimensions of Intentionality and the subsequent use of scripts to define a directedness extension that captures the narrowness of Intentional states; and (c) my treatment of error based on the aboutness-directedness distinction.

It occurred to me that scripts can be viewed as both Dretskean neural configurations and the concepts we possess, and I didn’t have to stray too far from Dretske’s own explanation of how this might be. Dretske argued that specific neural structures in the brain become concepts with content when: (i) they come to represent or indicate (i.e., react to in a specific way) certain things or states of affairs in the external world; (ii) this representative or indicatory function derives from the way the neural structures in question have been developed and used by the system of which they are a part—thus demonstrating original or intrinsic (as opposed to derived) Intentionality; and (iii) when these structures connect to and influence other such structures in a holistic way. Only when these three conditions obtain do otherwise syntactically-driven brain circuitry acquire content, and so become concepts in the full sense of the word. In this way, Dretske thinks that he has naturalized Intentionality.

My decision to link scripts to Dretske’s explanatory apparatus was motivated by more than just an inference to the best explanation—i.e., that scripts explain all sorts of cognitive phenomena. More than this, I recognized that the merger can give us leverage on the context in which acts of cognition take place and on the perspective of the thinker. The second of these goes to the heart of Intentionality and the related internalism-externalism debate: attempts to provide an account of thought without taking seriously the aspectual shape of Intentionality have always seemed to me to be silly demonstrations of wishful thinking. But then scripts came along and, together with Dretske’s information processing approach, showed precisely how to fix the contexts of mental acts and determine how things must seem from a discriminatorily limited first-person point of view: the keystone ideas behind the former are default-values and script-based inferences; the keystone ideas behind the latter are directedness extension and equivalence class.

In chapter I, I set out my agenda. After outlining the significance of Intentionality, I considered Intentional instrumentalism—the view that Intentional talk is just a convenient way of predicting an organism’s behaviour, but otherwise not to be taken literally. Instrumentalism is an example of a broadly dismissive stance toward the mental, a stance which nevertheless helps to clarify our options: we can either accept psychological irrealism and make ourselves comfortable with its consequences, or we can reject it and start our work from there. My sentiments incline me
Chapter VIII  PostScript

toward the latter course. The best defence for folk psychology will come in the form of a good working theory, and that will also give us reason to do away with instrumentalism and its kin.

In the second part of Chapter I, I briefly described two rival conceptions of Intentionality—internalism (or individualism) and externalism (or anti-individualism). This was followed by a statement of the position I wished to defend. It was there that the terms aboutness and directedness were used for the first time.

A longer, more precise definition of “Intentionality” was undertaken in chapter II. Four Intentional dimensions were identified: aboutness, directedness, aspectual (or perspectival) shape, and misrepresentation (or error). These elements have been much discussed in the literature, and making sense of them seems an important ingredient of any new proposal. I saw a purely externalist approach to the problem as both short-sighted and inadequate, not merely because it fails to accommodate the aspectual dimension of mental content, but because, convenience notwithstanding, it is not free to be so negligent. We are not at liberty to dismiss the “organismic contribution” to the content of thought, no matter how difficult it might be to formulate this contribution. The correct procedure is to relegate external influences to Intentional aboutness and then try to find a way of capturing the agent’s side of the equation.

But first we needed a closer look at internalism and its rival. In chapter III, I described the background against which externalism flourished and considered three types of anti-individualist arguments: those based on the physical furniture of the environment, those based on sociolinguistic conventions, and those emphasizing the causal-evolutionary history of persons. An attempt to trace the consequences of an externalist psychology was then made, followed by a discussion of individualism—in particular, the theories of early Jerry Fodor, John Searle, and Gabriel Segal.

I became acquainted with script theory quite by chance, and immediately saw in it the potential I was looking for. In chapter IV, I proposed to establish the foundations of script theory, starting with Conceptual Dependency and moving on to a description of the multiple long-term memory systems that house episodic, instrumental, personal, and definitional scripts. Of these four types, the last is by far the most elaborate, with divisions for animate and inanimate objects and a further division of animate objects into animals (pets, non-pets) and persons (historical figures, public personalities, passing acquaintances, friends, family relations, and perhaps even some kind of self-concept). I ended chapter IV with an outline of the basic structure of all scripts: headers, primes, tracks, frames, and default-values.

In chapter V, I focused on the evidence for script theory—in particular, computer task modelling and psychological support gathered from numerous experimental studies. The question of how scripts are physically implemented in the brain was the topic of chapter VI. It was there that Dretske’s information-processing account of mind really showed its versatility. Scripts were described as networks of neural structures which have been shaped—by evolution as well as the local learning history of the individual—to perform specific indicatory functions. This paved the way for explicating priming in terms of the Hebb-Marr theory of autoassociation and proposing LASCH as an explanation of a number of phenomena: priming by description, conceptual manipulation, cognitive holism, indirect learning, imagination, and the difference between human beings and animals incapable of natural-language communication.

Chapter VII was the place of reckoning. There I probed the extent to which the theory of scripts accounted for the Intentional dimensions discussed in chapter II. Aboutness was conceived evolutionarily but non-teleologically, and directedness was constructed with the help of the
mathematical notion of an \textit{equivalence class}. Aspectual shape was shown to be a function of the kind of information that a script provides and, finally, an account of error was given by comparing an agent's $E_a(S)$ and $E_d(S)$. Among the types of error revealed by this method are overgeneralizations, misprimes, overspecification, misidentification, mirages, optical illusions, and hallucinations. A different treatment for misrepresenting hypotheticals had to be provided. I also outlined the consequences of script theory to externalism. Two problems with causal-evolutionary approaches were raised: functional indeterminacy and the psychological non-status of Swampman. A final warning against radical internalism was sounded.

8.2 Where Shall We Go?

My work is not meant to offer a comprehensive theory of Intentionality. Indeed, I have doubts about whether such a theory can ever be provided, since Intentional phenomena do not appear to be homogeneous. There are many Intentional states that fall outside the boundary of script theory, and about these the approach I have advocated has nothing very interesting to say. The truth is, there \textit{are} limitations to schema-based accounts of cognitive processing. Marvin Minsky claims the following in regard to his frame theory:

"Schematic" thinking, based on matching complicated situations against stereotyped frame structures, must be inadequate for some aspects of mental activity. Obviously mature people can to some extent think about, as well as use their own representations. Let us speculatively interpret "formal operations" as processes that can examine and criticize our earlier representations (be they frame-like or whatever). With these we can begin to build up new structures to correspond to "representations of representations." I have no idea what role frame systems might play in these more complex activities. (1975:230)

Minsky believes that, whereas frame theory can help to explain first-order representations—what Piaget and Barbel (1971) called "concrete operations"—it may not be forthcoming when it comes to explaining second-order representations—what Piaget and Barbel called "formal operations."

I am not sure whether the same is true of scripts, but I hope it is clear that script theory can cover a lot of ground when it comes to Intentional mental activity. I am also confident that the same approach can illuminate other areas of interest in philosophy and psychology. Moral conduct is one such field. Do we have generic moral patterns that are stored in memory and "activated" to guide our behaviour whenever we enter the contextually appropriate conditions? We know that this \textit{is} the case in some social situations—vis, RESTAURANTes—but to what extent do scripts apply to ethics?

The question is at least worth investigating. We would not use scripts to \textit{construct} a general moral theory or devise any set of normative rules, but we may plausibly use them to get a better grip on a particular agent's \textit{moral psychology}. Why did Delilah become appalled when she thought that Samson stole ten dollars to buy cigarettes? Why did she then show more compassion when she realized that the money was used to buy milk for his children? What is it about Delilah's concept of "steal"—or her "steal" \textit{moral script}, if you like—that might explain these different reactions? In asking these questions, we have moved beyond the concrete (or nearly so) to the purely abstract. We are no longer talking about specifiable objects or roles or relations; we are talking about convictions and ideals. What inclines me toward believing that scripts can handle the more abstract issues reasonably well is that we are \textit{still} dealing with an agent's perspective, with the way the agent sees his or her world. Part of my project has been to show that script theory is apt at uncovering this elusive first-person point of view.
But it may be possible to move beyond the perspective of just the individual doer. How one group of people behaves towards another is often a product of the manner in which the second group is perceived by the first. It often happens that entire societies are beguiled by the same stereotypes, and when these stereotypes become derisive, disaster looms. The horrors of Nazi Germany were the example used in the text. It was an instance where the entire moral fabric of a nation was manipulated through a massive and persistent campaign of disinformation to better fit a government's particular sociopolitical ambitions. It isn't the viewpoint of a single individual that matters here, but the common viewpoint of an entire collective. How was this viewpoint created and how might it explain the actions to which it gave genesis? The potential use of scripts in this case straddles the fields of psychology and sociology.

Schank and Abelson mentioned a more clinical use of script theory. They suggested an approach whereby psychologists may probe the personal scripts (scripts,) of patients suffering from various kinds of neuroses, especially when these scripts are thought to represent distorted states of affairs or when they become activated inappropriately. The issue, once again, is the perspective of the patient and how this perspective interferes with normal socio-cognitive functioning.

These are just some of the ways script theory might be extended. In all cases, the scriptal approach—though not guaranteed to work any better than the incumbent strategies—represents a novel angle on old familiar problems, and this is enough to recommend further investigation. I have no doubt other applications are also possible. But my concern in this dissertation has been to show how scripts can help us understand the Intentionality of thought a little better. What I have presented is but a humble beginning. But even this much, in the complex and fickle world of cognitive science, is not insubstantial. We must learn to walk before we can run, and philosophers ought to do more walking.

Here is where I end my stroll.
Bibliography


Bibliography


Bibliography


