The Psychophysiology of Everyday Life

Philip Gerrans
Philosophy Program, Research School of Social Sciences
Australian National University

A dissertation submitted for the degree
Doctor of Philosophy
of the Australian National University
November 1995
Declaration

This dissertation is my own work, except where otherwise acknowledged.

Philip Gerrans.
Acknowledgements

My major debt is to Philip Pettit whom I would like to thank for many patient discussions which helped me understand the ultimate form my argument should take. Frank Jackson's work exemplifies an ideal I have tried to follow here, sophisticated argument in transparent prose. Karen Neander's expertise in cognitive science and her detailed comments improved the final draft immeasurably.

I am lucky to have been accepted into the PhD program at the RSSS. Almost all I know about philosophy I have learnt here (which thought may or may not frighten Frank and Philip) not only from colleagues but from the extended network of RSSS associates. Among those who had a specific influence on this thesis are: Tim van Gelder, who interested me in the empirical research on developmental psychology; Kevin Mulligan; Peter Lamarque, who quickened my interest in the philosophical aspects of psychopathology; Kim Sterelny; Simon Blackburn and Dan Sperber. Finally, and most importantly, I must thank Pauline Gerrans, whose immaculate proofreading was the least of her contributions to my work over the last four years.
Contents

Acknowledgements

Chapter One

1.1. The concepts of Intentional Psychology.
1.2. Scepticism about other Minds.
1.3. Intentionality and the Propositional Attitudes.
1.4. Intentional Psychology as Folk Theory.
1.5. Dennett on Pattern Recognition.
1.6. Logical behaviourism and the Dispositional Account of Mental States.
1.7. Ryle's Attack on Theory.
1.8. Theory and Structure.

Chapter Two

2.1. Tacit Knowledge.
2.1.1. Ryle on Knowhow.
2.1.2. The Regress Argument.
2.1.3. The Decomposition Argument.
2.1.4. The Neutrality Thesis.
2.2. Salience and Intentional Activity.
2.2.1. The Link Between Perception and Action.
2.2.2. Lebenswelt or Conceptual Scheme.
2.3. Peripheral Theory and Practical Knowledge.
2.4. Practical Knowledge.
2.4.1. Aristotle on Learning to be Good.
2.5. Seeing the Soul. Wittgenstein on Other Minds.
Chapter Three

3.1. The Child's Acquisition of Language.

3.2. The Chimpanzee Theory of Mind.

3.3. The Child's Theory of Mind.

3.3.1. CTM and ToMM: Modularity.

3.3.2. The False Belief Test.

3.3.3. Further Evidence for ToMM?

3.4. Theory of Mind. Constructed or Innate.

3.4.1. Early Stages in the Acquisition of a Theory of Mind?

3.4.2. The Understanding of Opacity.

3.5. Precursor States.

Chapter Four


4.1.1. The autistic triad.

4.2. Evidence for a Theory of Mind.

4.2.1. The Metarepresentation Hypothesis.

4.2.2. Metarepresentation and Opacity.

4.2.3. A Dispositional interpretation of the Evidence for Metarepresentation.

4.3. Theory of Mind and Behaviourism.


Chapter Five

5.1. Elements of Context.
5.2. Emotion and Cognition.

5.2.1. The Role of Emotion in Context Construction.

5.3. Emotion and Somatic Marking.

5.3.1. Evidence for Somatic Marking.

5.4. Context Construction and Somatic Marking:

5.5. Philosophical connections.

Chapter Six

6.1. Anthropology and Intentional Interpretation.

6.2. Evur amongst the Fang.

6.3. Anthropology and Psychology.

6.3.1. ToMM and the Rationality of Cultural Beliefs.

6.4. Bourdieu on Rule Following.

6.5. Anthropology and Psychology

Conclusion
Introduction

This thesis discusses a classical philosophical problem, our knowledge of other minds, and a classic analysis of that problem, that of Gilbert Ryle as a framework for the interpretation of current work in developmental psychology which deals with the young child's acquisition of psychological concepts. The most recent research into the development of the ability to understand others, in normal, Down's syndrome and autistic children, is framed by the assumption that our understanding of each other depends on tacit knowledge of a psychological theory. The very young child acquires this theory by trying out and discarding a series of predecessor theories until, by the age of four, she arrives at the adult understanding of other minds whose fundamental concepts, belief and desire, are understood as causes of behaviour.

The purpose of this thesis is to reinterpret the evidence on which this explanation relies and offer a different account of the development of mind-reading ability in children, and to extend that account to the general philosophical problem of understanding other minds. I take as a starting point for my account Ryle's famous analysis of a mental state, belief:

1 (Ryle, 1949 p135)
Ryle here makes a conceptual claim: that when we attribute mental states to others we are describing an integrated pattern of dispositions. He is often interpreted as identifying mental states with "bare" dispositions, or behaviours conceived of as outputs of a system. Whether or not this is correct, my interest in Ryle is not in either of these interpretations of his analysis of mental concepts, but in his claim that, in using mentalistic language we are not stating hypotheses about the nature of the entities which cause behaviour, we are using the resources of language to describe how the dispositions of an intelligent agent are related: they "hang together on the same propositional hook" because their relations parallel the logical relations among the sentences we use to interpret each other's behaviour.

This conceptual claim must be distinguished from an empirical issue: what causes us to have the interrelated dispositions which attract interpretation in intentional terms. This thesis is about both issues: on the conceptual front it endorses a Rylean view that the interpretation of behaviour, using mentalistic terminology, integrates a range of abilities and capacities of human agents, without purporting to state any hypothesis about the nature of the processes on which their integration depends. This perspective on the conceptual issue opens up a new perspective on the empirical issue. Recent research in the area has postulated a modular capacity of the human brain which represents a theory of human psychology, in much the same way as Chomskian linguists postulate a modular capacity of the brain which represents grammatical principles. I argue against this hypothesis and the interpretation of data on which it relies.

I develop a model of intentional psychology based on the data from autism. Autism is a developmental disorder which affects the ability to understand and relate to others: a failure to develop the ability to read minds. A satisfactory explanation of the difference between autistic and normal children should shed light on the nature of the mental representations involved in acquiring psychological concepts and thus on the relation between the conceptual and
empirical claims. The model I develop relies on the somatic marker hypothesis of Damasio which has two key elements. The first is that the mental representations involved in social understanding depend crucially on the integration of conscious or unconscious representations of emotional state. The second is that there is no site at which this integration takes place, the relevant mental representations are represented in parallel, not routed through a single module.

What separates the autistic and normal child is not that the latter has a modular capacity for the representation of psychological concepts as a theory, a set of propositions which articulates some domain in terms of fundamental underlying principles. Rather, the latter has intact a mechanism for integrating a variety of cognitive and emotional representations which functions in the way Damasio describes. This is the route to that conclusion.

**Chapter one** discusses the philosophical problem of other minds and locates the Rylean and the neo Quinean (a generic term for those who believe that our understanding of others depends on mastery of a theory of other minds) positions as responses to Cartesian scepticism about other minds. Ryle and the neo Quineans are shown to share foundational assumptions. The first is a hostility to Cartesian accounts which identify states of mind with immaterial introspectibilia. Ryle and the neo Quineans both give accounts of mind which locate it firmly in the physical world. Secondly, they share a conception of mental states as intentional, and as attitudes to propositions.

The **second** chapter examines the Rylean alternative on both the conceptual and the empirical fronts. On the conceptual front a distinction is made between theoretical knowledge which depends on explicit representation of theories and practical knowledge of the truths of a theory, which an agent can be said to possess in virtue of a set of dispositions. One can articulate an agent's practical knowledge, postulate a set of concepts known by the agent which explain her behaviour, without being committed to any hypothesis about the nature of the representations
on which that knowledge depends. The neo Quinean can thus be seen to be stating a specific empirical claim about the nature of the mental representations which underlie practical knowledge of the truths of intentional psychology.

In the third chapter, I examine evidence for a neo Quinean claim based on developmental psychology. I show, however, that the evidence only supports the claim that children acquire practical knowledge of the truths of intentional psychology, rather than supporting a specific, empirical hypothesis that the relevant knowledge depends on the functioning of a modular capacity of the mind which represents a series of theories about the causes of behaviour.

In the fourth chapter I make use of Sperber and Wilson's idea that the communication of meaning depends on the ability to mentally represent a context, the background information necessary to interpret utterance and gesture. My suggestion in this chapter is twofold. Firstly that interpreting others depends on the construction of a context which includes psychological elements like mental states, emotions and moods. Secondly that the construction of a context depends on a range of abilities and capacities, cognitive and non-cognitive. I examine the evidence from autism and argue that, while the autistic child cannot construct a context, the relevant deficit is not the failure of a modular capacity for theory-driven inference. I argue that the primary deficit in autism is the ability to integrate affective response with the representations of other psychological elements of a context.

In the fifth chapter I examine the empirical question. If my hypothesis about autism is correct, what does it tell us about the representations on which intentional understanding depends? I make use of Damasio's work on somatic markers to develop an account of context construction which explains both the autistic deficits and the normal case. That account treats intentional interpretation and social understanding as depending on a congeries of representations entertained in parallel. The resultant dispositions count as practical knowledge of
the truths of propositional attitude psychology, but there is no site at which a theory which states those truths and organises them via a set of unifying principles is represented.

The sixth and final chapter shows how the abilities developed by young children are applied as adults in the interpretation of word and deed in the wider social world. En route I discuss specific theories of acculturation and show how they are informed by Rylean or neo Quinean assumptions at either the empirical or conceptual level. One theorist explicitly aware of this is Sperber, who treats the acquisition of culture as a species of context construction and therefore, as relying on the same mental mechanisms as the acquisition of intentional psychology. While I rely on his account of context construction and agree with his way of stating the philosophical issues, we inevitably part company at the empirical level since Sperber endorses the view that the attribution of mental states depends on the functioning of a modular capacity for theory construction. In this last chapter I restate my principal thesis: the knowledge on which we rely in constructing a context for the interpretation of others depends on a diverse range of representations whose integration is not based on the representation of a theory which unifies that knowledge via the application of general explanatory hypotheses.
Chapter One. Understanding Other Minds

Precis

1.1. The concepts of Intentional Psychology. Understanding the use of mental state terms in the practice of everyday psychological explanation. Are they used to interpret and explain behaviour without stating any hypothesis as to the causal origin of behaviour, a basically Rylean position, or do they, as neo Quineans claim, state such hypotheses?

1.2. Scepticism about Other Minds. The neo-Quinean and the Rylean claim both shown to be responses to Cartesian Scepticism. The former treats mentalistic terms as predicates in a scientific explanatory theory, the latter as descriptions of an agent’s dispositional profile. In either case there is no essential mystery about the mind.

1.3. Intentionality and the Propositional Attitudes. Both accounts treat mental states as intentional and as attitudes to propositions. They share the feature of attempting to account for the meaningfulness of mental states in terms of the use of propositions to interpret and explain behaviour.

1.4. Intentional Psychology as Folk Theory. Arguments for the neo-Quinean claim that intentional psychology is a theory about the nature of the brain states which cause behaviour. The arguments from structural isomorphism and explanatory utility.

1.5. Dennett on Pattern Recognition. An alternative view of the way concepts can explain and interpret. Intentional concepts make certain behavioural patterns salient to us without implying any hypothesis about the way those patterns are brought about.

1.6. Logical behaviourism and the Dispositional Account of Mental States. Ryle’s treatment of mental state concepts as dispositions. The distinction between bare and categorical dispositions. Whether psychological explanations state laws. Two types of functionalism, commonsense functionalism which is not committed to the nomic view, and psychofunctionalism, the classic neo-Quinean argument that psychological explanations state laws relating brain states (or functional states of brains) and behaviour. Two ways of conceiving of behaviour, the first congenial to Ryle as a pattern of activity. The second, congenial to functionalism, as the output of a system.

1.7. Ryle’s Attack on Theory. With these distinctions in place, Ryle’s argument that psychological explanations do not state laws is examined. The neo-Quinean position, that psychological explanations state causal hypotheses, is shown to be motivated by the same anti Cartesian considerations which underpin Ryle’s argument.

1.8. Theory and Structure. Ryle and the neo Quinean positions shown to lie on a continuum, from Wittgenstein to modern Language of Thought hypotheses, of positions on the degree of internal structure we need to postulate in order to explain behaviour.
1.1 The Concepts of Intentional Psychology.

How do we know what, or if, other people are thinking? This is, of course, an issue in epistemology, a sceptical problem, but an answer to the sceptical question also raises issues about the nature of mental states. Metaphysical and epistemological issues in the philosophy of mind are not easily pried apart. To see this consider, briefly, three theoretical responses to the question, which I shall call Cartesian, Rylean and Neo Quinean.

The Cartesian thinks that we know what we ourselves are thinking by introspecting, and that if we know what others are thinking it must be on the basis of their reports or an inference based on an analogy between themselves and us. Different versions of the Cartesian argument draw a similar conclusion about the nature of these introspectibilia: that they are states of consciousness which do not depend for their existence on states of brain or body. Immaterial souls, for example, can have mental states. So a Cartesian epistemology is linked to a dualist metaphysic.

The neo Quinean response is based on an objection to this metaphysic: it claims that minds and their states have causal consequences in the physical world and must, therefore, be explicable in terms of the theories present or future about the nature of that world. On this view we know about mental states in the same way we know about many unobserved causes, by making inferences from observable evidence. Epistemologically and metaphysically the understanding of other minds is a project like the understanding of chemical valency or genetics, we rely on a theory of the domain in question to make inferences about the nature of the entities which cause observable regularities in behaviour. In the case of mental states the evidence is the behaviour of organisms like ourselves, the unobserved causes of which are mental states like belief and desire.

The Rylean position is an alternative to neo Quinean anti Cartesianism. It claims that mental states, like beliefs, desires, hopes and fears are dispositions to
Chapter 1  Understanding Other Minds

behave in certain ways. Ryle’s example, which will recur as a refrain throughout this thesis, is of a belief that the ice is thin.

to believe that the ice is thin is to be unhesitant in telling oneself and others that it is thin, in acquiescing in other people’s assertions to that effect, in objecting to statements to the contrary, in drawing consequences from the original proposition and so forth. But it is also to be prone to skate warily, to shudder, to dwell in imagination on possible disasters, and to warn other skaters. It is not only a propensity to make certain theoretical moves, but to make certain executive and imaginative moves, as well as to have certain feelings.\(^1\)

The argument between Ryleans/Wittgensteinians and scientifically-minded philosophers like the neo-Quinean is an old one which has recently resurfaced as philosophy confronts developments in cognitive and neuro-science. The debate takes for granted a physicalist ontology and the notion that the behaviour classified and interpreted by mental states is the causal outcome of neurobiological processes within the organism. The debate then takes place on two fronts. The first is an argument over our use of mental state terms. Whether, as the neo Quinean claims, we use them to state hypotheses about the nature of causally efficacious neurobiological states, or as the Rylean claims, we are talking about abilities and capacities however realised.

I claim that in using mentalistic vocabulary we are articulating the relations between an agent’s multi tracked dispositions rather than stating hypotheses about the causes of behaviour. At this point a crucial issue arises: a neo Quinean might grant that the Rylean view captures more of the intuitive meaning of mental state terms than the neo Quinean analysis. She might agree that what we are talking about when we use the idioms of intentional psychology are the integrated dispositions of our colleagues but she may still maintain that agents exhibit those dispositions as the result of the operation of a

\(^1\) (Ryle 1949 p135)
module which represents a psychological theory. Dan Sperber makes a version of this move when he says:

We might be tempted to say then, as many philosophers have, e.g. Ryle (1949) that a belief is a disposition to assent to a proposition. As psychologists, however, we will want to go deeper: what kind of mental states might bring about such a disposition? ²

By mental state Sperber here means a state of the brain which causes behaviour. So there is another front on which the Rylean/neo Quinean debate takes place. The debate at the conceptual level is over whether the terms of intentional psychology refer to dispositions or state hypotheses about the causes of behaviour. At the empirical level, which is Sperber's primary interest, there is the question of the nature of the actual mental representations on which dispositions depend. Although I think the Rylean position captures more of the intuitive meaning of mental state terms than the neo Quinean, I am also interested in the empirical question. So, unlike Ryle, I do not regard the matter as closed once we take a position on the the conceptual issue. My main interest in Ryle is, in fact, the use to which his conceptual claim can be put in the generation of interesting empirical theories.

I claim that the Rylean perspective, which is neutral about the nature of the causal processes which underlie behaviour, is a better theoretical starting point for the investigation of mental representations on which the dispositions classified by intentional psychology depend. I develop these claims by reinterpretating the evidence for recent theories in developmental psychology which rest on a specifically neo Quinean conception of the nature of mental states. These accounts claim that the child develops her understanding of others by learning a theory which postulates unobserved states of mind, with specific properties, in order to explain others', and her own, behaviour. The theory acquired is the theory of propositional attitude, or intentional, psychology. The

²(Sperber 1990p32)
child does this, it is alleged, via the activation of a modular capacity of the mind, in much the same way as the acquisition of language depends on dedicated neural circuitry which represents grammatical principles. The interest of the Rylean alternative is that if we see the child’s developing understanding of other minds as depending on the integration of a range of capacities, rather than the activation of a discrete theoretical module, our understanding will be enhanced. In support of this claim I give a new explanation of the phenomenon of autism, a paradigmatic case of failure to understand other minds, in chapter four, arguing that what the autistic child lacks is the ability to focus a range of abilities and capacities in the project of understanding others, rather than a discrete ability to theorise about the nature of the causes of another's behaviour.

1.2 Scepticism about other Minds

Any explanation of an agent’s behaviour which invokes her mental states confronts the venerable philosophical problem of scepticism about other minds. How do we know what others are thinking? or if they are thinking at all

The neo Quinean and Rylean positions are accounts of intentional psychology which can be understood as responses to classical scepticism about other minds. They are answers to the question "how do we know what (or if ) others are thinking ? " As well as this anti sceptical bent they share other theoretical features: having rejected a Cartesianism which treats first person access as the essential feature of mental states, both Ryleans and neo Quineans treat the intentionality of mental states and their availability for articulation in language, via the fact that mental states represent propositions, as crucial explananda for their accounts. The rest of this chapter explains how Ryle and the neo-Quineans account for intentionality and propositionality within their respective anti sceptical, anti Cartesian frameworks.
The interest of these positions is that, in some respects, the autistic child never develops a satisfactory account of others' mental states, she is someone for whom the sceptical question can never be allayed. Because of their inability to understand other minds, even the highest functioning autistics are never secure in social situations which require them to interpret others' intentions and feelings and respond appropriately. If a satisfactory account of the difference between autistics and the normal person can be given, it should include an account of how it is that we normally acquire our knowledge of what others are thinking and feeling. The main task of the later chapters of this thesis is, as I said, to re-evaluate a neo-Quinean claim that the difference between the autistic and the normal child is that the latter solves the sceptical problem by developing a modular capacity for the generation of empirical theories about the causes of others' behaviour. The task of the first chapter is to show how the sceptical problem arises within a Cartesian framework and how the Rylean and neo Quineans respond.

On a traditional account of the epistemology of the mental, which distinguishes first person access to the content of one's thoughts from the type of access which third persons have, the problem arises in this way. If I consciously attend to my own thoughts then I automatically know what I am thinking. Other people, however, have access to the contents of my mind in a different way, they are dependent on my sincere reports or the evidence of my behaviour when they try and ascertain what I am thinking. This phenomenon of privileged first person access to the contents of thought, combined with the further view that incorrigible introspective access is the distinguishing feature of the mental, automatically generates semantic and epistemic scepticism about intentional explanation.

This phenomenon of privileged first person access to the contents of thought, combined with the further view that incorrigible introspective access is the distinguishing feature of the mental, automatically generates semantic and epistemic scepticism about intentional explanation.

The epistemic problem is quite simple: when I report that someone else believes that p, I am not reporting that fact on the basis of my experience of the mental state in question. Thus, because first person access is the only way to

---

3 The classic paper here is Rorty's 'Incorrigibility as a Mark of the Mental' (Rorty 1970)
gain knowledge, as opposed to hypothetical speculation, about thoughts, I can never know, but only conjecture, what another is thinking. The possibility that other people do not have minds or are not actually thinking the thoughts we attribute them on the basis of their public performances, cannot ever be ruled out if the only access to minds is first person conscious awareness.

Semantic scepticism, based on characterising mental states as those to which we have incorrigible privileged access, arises in this way. Meaningful statements require that there be intersubjective standards against which they can be checked. If I say "the play starts at 7:30" then I am right or wrong in virtue of whether or not the play starts at 7:30. The meaning of my statement, composed of its semantic constituents, allows someone who understands the publicly determinable meaning of those constituents and the syntactic rules of English to determine whether I have spoken truly or not.

In the normal case, questions about the meaning of a term or a sentence can be settled by a dictionary, translation manual or by perceptual confirmation. If someone avers that "dogs are feline" or that "sofas are religious altars" then we can find out what she means, firstly by getting her to point to, or describe, a dog or a sofa and then asking her to look up "altar" and "feline" in the OED. However, if mental state terms refer to private experiences, accessible only in introspection, this procedure will not work. Consider two people who report their beliefs that "the sunsets in California are more fiery than those of Arizona". If they are both making this report on the

---

4 How to understand knowledge gained in this way is difficult. Most of the discussion focuses on sensations of, for example, colour experience, taste or pain because it seems, there is no doubt that my knowledge of such mental states is gained in a different way than that of outside observers. If, one couples the view that such knowledge is of "irreducible phenomenal information" (the term is David Lewis') with the idea that phenomenal experience is essential to the having of such states (as does Kripke) then one always has to allow the possibility that someone who seems to be in pain, seeing red etc may only be behaving as if they are in the relevant mental state. However to get the sceptical argument started one only needs the premises (however that is decomposed under analysis) that I know about my mental states in a way which others do not and that privileged access is the distinguishing mark of mental states. The second premise is the one attacked by materialists such as Sellars, (Sellars 1956) who seek to show that our first person knowledge of mental states does not mark them off in any interesting way from other natural phenomena and hence insulate them from scientific investigation. Churchland's neo Quinean argument is a direct descendant from Sellars' attack on the privacy of mental experience.
basis of introspective access to their mental experience there can be no criteria
to which these two people, or even earlier and later stages of the same self,
who report the same mental event, can resort, to determine whether or not they
are talking about the same things.

Thus every time we attribute mental states, to other people or ourselves,
the sceptical problem arises in one or other form. How do we know that our
attribution of a mental state is really what someone else is thinking? How do
we know that we are all talking about the same things when we describe
mental states?

One approach is to brush off scepticism about other minds in the same
way one brushes off global scepticism. The possibility that other people might
not have minds or might be behaving as they do for reasons which are utterly
inaccessible to us is no more interesting than the possibility that we might be
brains in vats or dreaming. However, someone who takes this approach is still
left with the problem of giving an account of the domain of knowledge under
sceptical attack. This problem becomes particularly acute in the realm of the
mind precisely because the sceptical problem arises in a number of forms
which challenge us to develop an account of the nature of mental states and our
knowledge of them. Consider just three examples. A chimpanzee sees a second
chimpanzee enter the cage. She stays still until the other chimpanzee leaves.
She looks around and then extracts a banana from under a rug and eats it. The
second chimpanzee, who had hidden outside the cage, then returns, uncovers
the cache of bananas and eats them. This type of scenario, endlessly repeated in
chimpanzee colonies, invites explanation in these terms. The first chimpanzee
did not want the second to know where the bananas were so she hid them. The
second pretended to be deceived and then returned to eat the bananas. Do
chimpanzees really have concepts of these italicized mental states and plan and
execute their actions accordingly?

Or imagine that a psychoanalyst tells you that your failure to complete a
thesis on time is the result of a desire to castrate your father, or your father’s
desire to castrate you, (take your pick). Or imagine that an anthropologist, explains an instance of behaviour as the outcome of a set of beliefs in nocturnal occult forces whose efficacy extends only to the boundary of the forest. On what basis do psychoanalysts and anthropologists reach these conclusions? Surely one is justified in asking "How do you know?" in cases such as this.

Or imagine that the nanotechnological ambitions of cognitive scientists are realised and that it becomes possible to replace each cell in your brain one by one by a microscopic piece of ceramic which conducts electrical current according to the strength of electrical excitation in adjacent cells. Eventually all your neural circuitry is replaced and you continue to speak and act as before. The cognitive scientist says, firstly, that all your beliefs, memories, hopes and fears, are intact and, secondly, that the ceramic brain represents the propositions used to ascribe such mental states, such as "fears sharks", "believes that Sophie's World is a better work of philosophy than Theory of Justice" and so on. How does the cognitive scientist know this?

In all these cases, an answer to the epistemological questions would be unsatisfactory without an account of the mental states in question and their relation to the behaviour which invites their attribution. Just to say that the behaviour of the chimpanzee/shaman/cyborg is good evidence of the presence of mental states is inconclusive without an account of what a mental state is and how it connects up with that behaviour. Of course attributing mental states in these unfamiliar cases is just an extension of our everyday unreflective use of mentalistic vocabulary to explain behaviour, deliberate and reflect. However the marginal cases force us to reflect on the everyday, unreflective practice and ask "what justifies our ordinary use of mentalistic vocabulary?" To answer this question we need an account of the use of mentalistic vocabulary and likely occupants of the role of mental states within that practice.

It may turn out that our use of mentalistic vocabulary is not sufficiently univocal and precise to identify a unique occupant. In that case we have some
options. Abandon the normal use and redefine the concept more precisely to include only certain cases and exclude others. For example, include the normal human and the chimpanzee but exclude dogs and ceramic brains as bearers of mental states. Or we might include the ceramic brain and the human but exclude chimpanzees. Another option might be recognise metaphysical or epistemic limitations on our use of mentalistic terminology. In the metaphysical case it might be genuinely indeterminate whether nature has sharp boundaries. In the epistemic case it might be the case that nature has determinate boundaries which exceed our epistemic capacities. In either case, our use of mentalistic terms will be imprecise, either because minds come in degrees, or because we cannot detect the properties necessary to distinguish the minded from the non-minded.

Perhaps it will turn out that nothing has \textit{all} the properties our unreflective use of mental state terms implies. We may have been using a concept to interpret and explain our behaviour which we ourselves only partially satisfy. If, for instance, the concept of belief is tied to a standard of rationality which we do not ordinarily attain, then we humans will only be partial believers, less than fully rational. We may be the best deservers of the attribution of mentalistic vocabulary at present but perhaps other organisms and machines may turn out to be even better thinkers than us.\footnote{Stephen Stich's view in \textit{Fragmentation of Reason}. (Stich, 1989)}

Which, if any, of these options is appropriate depends on what we can ascertain about the everyday practice of explaining behaviour in terms of mentalistic vocabulary. Those explanations indicate that we expect mental states to have these minimal properties, they are intentional and they represent propositions. Once this is established the problem for the neo Quinean or the Rylean becomes a matter of giving an adequate account of intentionality and the mental representation of propositions.

\textit{1.3 Intentionality and Propositional Attitudes}
The explanation of actions invokes mental states which fall into two classes: those mental states which represent the world, like belief, knowledge and expectation, and those which, given our beliefs about the world, motivate us to act in that world. Desires, fears and aversions are such states. The two classes of mental state (for which Belief and Desire will henceforth serve as proxies) can be characterised in terms of the way they interact with the world. Beliefs aim at accurate representation of the world, thus they must be revised as new information comes to hand. Desires, on the other hand, are satisfied by changes in the state of the world. A true belief reflects the world as it is, a desire represents a state of the world as we would like it to be.

The everyday practice of action explanation invokes these two classes of mental states in order to allow us to recognise actions as falling under this general pattern:  

Where X is an agent, O is a state of affairs and A an action: If  
1. X desires O and has no other overriding desire.  
2. X believes that A is the best way to bring O about. Then,  
3. Ceteris Paribus, X does A.

Everyday explanation of behaviour subsumes instances of behaviour under a pattern established by typical connections between beliefs, desires and actions. If a friend sees an advertisement seeking donations to a charitable cause and gives some money, we attribute to her the desire to assist the needy. And vice versa. If we know that she is charitably inclined and we see her generous behaviour to a person or organisation, we can infer that she believes that the people need assistance.

---

6 This is a crude characterisation of the difference between beliefs and desires in terms of "direction of fit". For an extended discussion of the relevance of the distinction in explaining human action see (Smith 1987). (Humberstone 1992) gives a brief but thorough survey of the history of the distinction in the course of a sophisticated characterisation of what he terms telic and thetic attitudes, distinguished by the direction of conditionalisation between their controlling intentions and the world.

7 The pattern given here is an abridgement of the more detailed discussion of Churchland's characterisation of propositional attitude psychology by Pettit. (Pettit 1986p24)
Because this type of explanation recognises typical connections between beliefs, desires and actions it assumes a minimal standard of rationality: the production of actions designed to realise goals, given a set of beliefs about the way the world is and the best way to realise those goals. We narrow the range of possible explanations of behaviour to the one which cites the most relevant or likely beliefs and desires of the agent. This is why Anscombe said of intentional psychology "The interest of the account is that it describes an order which is there whenever actions are done with intentions." 8

It is important to remember that intentional psychology does not require that we pretend, for the purposes of interpreting others' behaviour, that they are ideally rational, just that their beliefs are about the way things stand in the world independently of them and that their actions aim at satisfying their desires. Most of anyone's beliefs are contained in an enormous list of uncontroversial truisms about their environment. Things like the belief that water quenches thirst, that there is water in the fridge, that the fridge is at the other side of the room and so on. We attribute patterns of beliefs and desires on the basis of large scale patterns of activity and in virtue of a total profile of, mainly platitudinous and tacit, beliefs and desires. Even though explanations tend to cite only one or two salient beliefs in explaining action, those attributions hold a large background of relevant beliefs and desires constant. The cases on which experimental psychology tends to focus are those, relatively rare, contexts where, as a result of explicit reasoning, we gain or revise beliefs one by one, but, as a matter of fact, any time we receive new information about the world our belief-desire profile tends to alter accordingly.9

8 (Anscombe 1957 p80)
9 Although it seems to be a psychological fact about humans that any such revision is local rather than global. We do not successfully revise every belief we have every time we gain some new information about the world. Rather, new information tends to affect only those beliefs stored in the immediate inferential vicinity. This is presumably why we can maintain contradictory beliefs until their incompatibility is demonstrated explicitly. The context in which beliefs are acquired or employed tends to insulate them from beliefs which are accessed in other contexts. It is this context relativity of belief acquisition which gives rise to the phenomenon of opacity, in its starkest form demonstrated by the predicament of Puzzling Pierre, who has acquired two contradictory beliefs about London 'that it is pretty' and 'that it is ugly' but cannot recognise that these beliefs are both about the same city because he has acquired them in different linguistic and geographical contexts. (Kripke 1979)
So our everyday explanation of action interprets behaviour in terms of the rational adjustment of actions to realise goals in a world as represented by the agent's beliefs. This is one aspect of the intentionality of mental states. Intentionality comes from the Latin *intendere* —"to incline towards" and captures the so called "directedness" of mental states, the fact that they are "about" the world as represented in the mind of the agent. Intentional states are not always directed to the concrete physical world, we can also think about worlds made up of fictions, possibilities and abstracta. As David Lewis puts it:

If I hear the patter of little feet around the house I expect Bruce. What I expect is a cat, a particular cat. If I heard such a patter in another house, I might expect a cat but no particular cat. What I expect then seems to be a Meinongian incomplete cat. ...

If we let surface grammar be our guide, the objects of expectation seem quite a miscellany. The same goes for belief, since expectation is one kind of belief. The same goes for desire: I could want Bruce, want a cat, but no particular cat, want winter, want stormy weather, want to shovel snow, want fatigue, or want that someday mankind will inhabit at least five planets. The same goes for other attitudes to the extent that they consist partly of beliefs and desires or lacks thereof. 10

In rationalising action we attribute the beliefs and desires which characterise the represented worlds, not the world as it is. Beliefs, for example, are not a straightforward reflection of the world, often they are false, based on misleading or out of date information or defective reasoning. David Lewis opens a tin of cat food whether or not his belief is satisfied by Bruce, a stray cat, a small dog or an auditory hallucination. However in explaining his action we cite the belief that "Bruce is in the hallway." Desires have the same property, one can wish for, or fear, states of the world which are unrealised like world peace or nuclear holocaust.

10 (Lewis 1979).
What then are the objects of these mental states, the things to which they are directed? As Lewis says, if we let surface grammar be our guide, they are a miscellany, some mental states are about things in the actual world, some are about things in other, non actual worlds. (For instance I believe that no unicorn has two horns.) Intentional psychology solves this problem by treating mental states as attitudes to propositions, the meanings of sentences. A belief is always a belief that \( p \), a desire, a desire that \( q \), an apprehension an apprehension of \( r \) where, \( p \), \( q \), and \( r \), specify the way the world must be if the belief is to be true, the desire satisfied, and the apprehension realised.

Treating mental states this way has two advantages. The first is that propositions stand in logical relations to each other. They may imply or contradict each other for instance. This allows the rationality of the connection between mental states and actions to be expressed in terms of the logical relations among the propositions we use to explain the actions. The second is that it allows us to express a feature of mental states which derives from their intentionality, the fact that they are not direct or transparent mirrors of reality.

Frege noticed that this inference is invalid.\(^{11}\)

\(^{11}\) It is plausible that mental state terms are subject to another, related logical restriction discussed at length by (Quine 1956) and (Kaplan 1985) The restriction is the ban on exportation, of quantifying into intensional contexts. In purely extensional contexts one can make the following inference:

\[
\text{Fa} \\
\exists x: x = a \land Fx
\]

In intensional contexts however it seems one cannot move from the fact of a belief that \( \text{Fa} \) to the existence of the object believed to have a certain property. This inference is banned:

\[
\text{S believes } \text{Fa} \\
\exists x: x = a \land \text{S believes, of } x, \text{ that it } \text{ is } F.
\]

The ban on exportation is controversial. It depends on one's view of singular thought, in particular of judgements made on the basis of perceptual demonstration which seem the most likely to license the inference. See (Burge 1977a, Burge 1977) for defence of the view that beliefs are individuated on the basis of environmental connections and that singular thoughts entertained on the basis of indexical presentations are primary in constructing a believer's conceptual repertoire. Belief \( \text{de re} \), which licenses this form of inference is therefore the foundation of intentional explanation.

For an opposing view, universalism, that belief is always \( \text{de dicto} \) see (Blackburn 1984). The \( \text{de dicto} \) view fits best with certain theories in cognitive science and perception to the effect that belief is always a matter of entertaining a certain mental representation which fits the world or not in view of its causal history. Dennett argues against \( \text{de re} \) belief in (Dennett 1982). His final view on the matter, and the one which is orthodoxy in cognitive science is: "the common assumption that it is not possible to make a judgement about an object without knowing what object you are making a judgement about – called Russell's Principle by Evans – must be abandoned. For all its intuitive, introspective appeal, it is a source of the lion's share of the theorist's misfortunes. (Dennett 1987p204)" In 'Quantifying In' Kaplan defends the view that \( \text{de dicto} \)
Chapter 1

Understanding Other Minds

Philip believes that the Evening Star is identical to the planet Venus.

The planet Venus is identical to the Morning Star.

C. Philip believes that the Evening Star is identical to the Morning Star.

All the singular terms refer to the same object, yet the inference does not go through. The logical formalisation of all sentences which express intentional attitudes runs into the same problem: substitution of a co-referring singular term into a sentence does not preserve the truth value of the whole sentence where the substitution takes place within the scope of a psychological verb.

The culprit is the intentional aspect of mental states. It is quite possible to believe that Venus is identical with the Morning Star without believing that Venus is identical to the Evening Star — because those expressions describe two different ways of thinking about Venus. The same holds true for other psychological attitudes. Someone might desire to have dinner with Edna Everage without any desire to have dinner with Barry Humphries, even though they are in fact the same human being. In this case the different desires are based in two different ways of thinking about the same human being. These different ways of thinking are expressed in intentional psychology by different propositions.

What is called propositional attitude psychology is thus a matter of understanding people by attributing to them thoughts, individuated by propositions, which explain their behaviour in terms of the inferential beliefs can sustain existential generalisation when they are sufficiently rich to give the believer cognitive acquaintance with their object. The fullest discussions of the logical problems associated with de re belief are in (Boer 1975).

12 Or so it is often conveniently assumed. Intentionality as Brentano originally used it referred to the way in which the mind can direct itself to objects, including nonexistents. The phenomenon Frege noticed, intensionality, is the failure of extensional substitutions in certain psychological contexts. This is usually explained in terms of the intentionality, in the Brentanian sense, of the psychological attitudes. Intentional content is expressed by a proposition, so that one cannot give the logical form of a sentence reporting a belief, without quantifying over an intension. See (Quine 1956)
connections among those propositions. Although Frege was doing logic, not psychology, he was taking over from intentional psychology a characterisation of mental states which individuated them by their inferential relations with each other. In other words, he assumed that intentional psychology describes mental states, whose conditions of satisfaction were individuated by propositions.13

Intentional psychology and propositional attitude psychology are thus interchangeable terms for our everyday practice of explaining action by attributing attitudes of belief, or desire, to propositions.

However this does not advance us on either the epistemological or metaphysical fronts until we know, firstly, what it is for a person to represent a particular proposition and, secondly, how it is that we know of a person that her actions are the result of representing a particular proposition. The neo

13 Actually, Fregean Arts des Gegebenseins (literally, the way an object has of being given in thought) are expressed by the Fregean Sense of a context free sentence, one which uses no indexicals or tensed expressions. Church translated Frege, using the term proposition to stand for a Fregean thought. For more on this point, and the way the equation of Fregean Thoughts with propositions has led to a slight distortion of his thought see (Burge 1979a) especially p407ff and note 10. In fact propositions play at least three, apparently incompatible, roles in a Fregean theory. Firstly, they express the agent’s perspective on the world. The way she recognises an object in thought, whether or not the object is physically present. Secondly, in their semantic role, propositions express a Fregean Thought, something like a definite description which describes the object of the thought in such a way that any two people who think a thought with that content are thereby guaranteed to be thinking about the same object. Thirdly, they function in their familiar rationalising role of explaining behaviour specifying the content of the beliefs and desires which make the behaviour intelligible. (Burge 1977) was the first to make this tripartite distinction, Salmon proposed a more refined taxonomy in (Salmon 1981). (Oppy1992) gives a brief, but detailed, discussion of the many logical and psychological roles which Fregean senses have been called upon to play. There are many who deny that a semantics of propositional attitude expressions requires us to quantify over something like Fregean senses. However these people are not denying that agents’ actions are explained by their way of conceiving of their environment. They are denying that those mental representations are directly specified by the meanings of the sentences of intentional explanation. Nonetheless they say that the content of the representations is still communicated by intentional explanation. The communication takes place pragmatically, by a process of Gricean implicature rather than directly through the semantics of the terms used in intentional explanation. This view originally derives from reflection on the problem indexicals and tensed sentences seem to pose for the Fregean theory and leads to the view that what is communicated by sentences is less a matter of their dictionary, or abstract meaning, than what the hearer imports from the context. See (Salmon 1986; Wettstein 1986; Schiffer 1992) for criticisms of the view that the semantics of propositional attitude terms require quantifying over Fregean senses. Perry has been developing a version of the view which accepts that the way a subject represents her environment explains action, linked to a strong Fodorian view of the nature of those representations as causally efficacious brain states with something like ‘propositional content. Nonetheless Perry wants to deny that the terms “belief” in the sentence “Sharon believes that p” refers to a Fodorian representation. (Perry 1989; Perry 1990; Perry 1991)
Quinean and the Rylean can be seen as giving two different anti Cartesian accounts of this issue.

Quine did not believe in propositions, he believed that the world was made up of physical objects and their relata described by the laws of physics. Thus he felt that propositional attitude vocabulary was pointless and misleading since it relied on propositions ("creatures of Darkness") to carry so much explanatory weight, while the laws of physics, or any science reducible to physics, made no mention of propositions. Thus the only hope for intentional psychology was to translate it into more austere scientific language. If no such translation were possible, then we might have to farewell intentional psychology.

If we are limning the true and ultimate structure of reality, the canonical scheme for us is the austere scheme that knows no quotation but direct quotation and no propositions but only the physical constitutions and behaviour of organisms.

He took the view that no such translation was possible, proposing instead that intentional idioms should be replaced, in the interests of science and logic, by the language of scientific psychology. He was thus an early eliminativist about, if not mental states, mentalistic vocabulary.

When Quine first suggested that intentional explanation should be thus reduced, the preferred theoretical language of scientific psychology in the United States was a behaviourist language of stimulus and response. There are now other flourishing research programs in the scientific explanation of behaviour. Neurologists, for example, attempt to explain why agents do the things they do in terms of information processing in the brain. Cognitive scientists attempt to model cognition in terms of computational architecture,

---

14 Actually Quine referred to Intensions as creatures of darkness but, in the spirit of Church and Goedel, I equate the Intensions with propositions in discussing Quine. Quine's general hostility to abstracta licenses the approach. (Actually Quines counted some abstracta, such as sets, but he was hostile to abstract objects individuated by meaning relations.)
15 (Quine 1960p221)
Classical, Connectionist or Dynamical. Neo Quineans are those who, like Quine before them, require that intentional explanations must be translated into the language of scientific psychology, now conceived, not in behaviourist, but in neuro or cognitive scientific terms. The first step in such a project must be the rendering of intentional psychology as a theory of the same domain as scientific psychology in order to commence the project of reduction to the more scientific theory or elimination in its favour.

1.4. Intentional Psychology as Folk Theory

A theory is a set of sentences which is articulated, comprehensive and cognitive within its domain. Articulated in the sense that each sentence plays its explanatory role as part of the network made up by the complete theory. Theories are comprehensive in that they aim at explanation of all the facts within the domain through their employment of fundamental explanatory principles and projectible concepts. Someone who understands the theory can, unlike someone who has merely seen a list of elements, determine whether new facts are within the domain of the theory and how they relate to the other facts which make up the theoretical domain. It is this requirement which separates theories from lists or descriptions. One could, for example, write down every possible grammatical English sentence, or the valence of every element, but, while that project might provide some useful information, it would not be an explanatory theory of English grammar or chemical valency. A theory displays, through its conceptual organisation, the structural organisation of its domain. This is why zoological and botanical classification are theoretical enterprises and lists of plants or animals are not. A taxonomy, unlike a list, is constructed according to principles which articulate the structure of the domain.

In this respect, theories are cognitive, carrying a commitment to revision if facts turn up which cannot be explained by the theory or if a better theory of the same facts turns up.
These requirements, that a theory be articulated, comprehensive and cognitive are minimal. They are not intended to obscure the differences between theories of literature, physics, linguistics, or psychoanalysis. Within their domain, theories have an important epistemic role. They provide criteria for belief formation and revision. Someone who understands a theory can use it to make predictions and interpretations. When a sociologist predicts that the crime rate will rise in an area of increasing unemployment or an economist predicts that interest rates will fall, they are doing so on the basis of sociological or economic theories. One can use theories to extend one's knowledge beyond the data from which the theory was compiled because theories are constructed around the generalisable properties of their domain.

Churchland regards intentional explanation as proto scientific theory, and physics, conducted against a background of scientific realism, as the paradigm of scientific theory. Thus he represents the inferential relations among propositions in intentional explanation as paralleling the mathematical relations among variables in physical theory. According to Churchland, the relations among both sets of abstracta have the same function: to direct us to the causally efficacious properties of the entities which are the domain of the relevant theory. Brain states in the case of intentional explanation, or, in the case of physics, properties of regions of space time.

On his view of intentional explanation, we postulate, on the basis of behavioural observation, laws which state the relations between mental states and behaviour. We then use those laws as premises which enable us, in combination with premises about observed behaviour, to deduce the properties of the mental states which caused that behaviour.

Churchland\textsuperscript{16} gives the following example of a scientific explanation:

All copper expands when heated
This bar is copper
This bar is heated

\textsuperscript{16} (Churchland 1988p57)
This bar expands

The first premise is an empirical hypothesis linking the properties mentioned in the other two. The second and third premises are empirical facts. The conclusion then follows deductively from the premises. It should be apparent the first premise is functioning as a covering law. It is able to do so because it is an empirically well confirmed theoretical hypothesis. Now consider the structure of intentional explanation:

\[ X \text{ desires } O \text{ and has no other overriding desire.} \]
\[ X \text{ believes that } A \text{ is the best way to bring } O \text{ about.} \]

*Ceteris Paribus, } X \text{ does } A.

One could easily rewrite this principle in the form of a *ceteris paribus* covering law: Rational agents act to satisfy their desires in accordance with their beliefs. The attribution, on the basis of behaviour, of the content of a belief then comes out, in a typical instance, as follows:

Rational agents act, *ceteris paribus*, so as to satisfy their desires according to their beliefs.

\[ S \text{ is a rational agent} \]
\[ S \text{ desired to see the play.} \]
\[ S \text{ went to the theatre at 7:30} \]
\[ S \text{ believed the play started at 7:30} \]

The structural isomorphism with the causal/covering law model of scientific explanation is apparent. Churchland claims that the conclusion is reached by the same route in both cases, via the employment of an empirical hypothesis, a putative law, as a first premise. The conclusion itself becomes an empirical hypothesis, available to be tested against the evidence. The results of
those tests affect the strength of the first premise, so the terms in empirical theories are interdefined by roles they play in the theory.

The structural argument has another strand, based on similarities between the role played by propositions in intentional explanation and the role of the numerical relations between abstract objects, like mass and velocity in physical theory. The numerical relations between abstract objects allow us to explain and predict physical phenomena in the same way as the inferential relations among propositions allow us to explain and predict human behaviour. For example, numerical relations between these abstract objects are exploited to enable us to construct explanatory generalisations such as: objects with mass \( x \) and velocity \( y \) take a force of \( z \) to arrest their motion. As a result, we can explain why a truck needs stronger brakes than a car. In the same way we say people who fear \( p \) desire, \( \text{ceteris paribus}, \sim p \). We then use this law of the relation between fears and desires to explain why someone, afraid of sharks, who believes that there may be a shark in the water does not want to go swimming. To say that someone refused to go in the water because they were afraid of sharks is to theorise about the motive for their behaviour. A trivial piece of theorising to be sure but one well licensed by the massive network of commonplace theory and mundane lore we use to predict, interpret and make generalisations about human action.

Apart from these structural similarities, there is the explanatory utility of treating intentional explanations as an empirical theory of mind. Treating it as an explanatory hypothesis about the laws which explain the relations between inner states and behaviour, systematises the solution to various problems in the philosophy of mind. In particular scepticism about other minds becomes no more or less worrying than scepticism about electrons or hypothesised genetic structures. On the neo Quinean view, when I say 'S wants to see Romeo and Juliet performed' I am referring to a state of S's brain which, in combination with the rest of her beliefs and desires, explains her actions. Of course there may be no such state, but this is not an occasion for global scepticism about
intentional explanation, it is merely an instance of the failure of a fallible inference, from observed behavioural evidence to unobserved causes. In general, we do not regard our theoretical judgements as being undermined by the possibility that they license the occasional false inference. Rather, our theoretically conditioned judgements stand or fall according to their explanatory utility, coherence with other theories in intersecting domains, and the accommodation of further data by the theory.

When a chemist says that "Oxygen has a valence of two" she is inferring, from the observed behaviour of oxygen in experiments, that it has a physical structure which explains its chemical bonding behaviour. Valency theory may be wrong, but as long as it continues to explain chemical behaviour and to fit with the development of theories in contiguous domains, there is no reason to dispense with it. It is a well confirmed and useful, though fallible, theory. If we treat intentional explanation in the same way, then our judgements about what other people believe and desire are just another example of fallible inference from observed behaviour to the unobserved causes thereof, namely their mental states, based in a well confirmed theory of behavioural explanation.

On these two grounds, Churchland concludes:

The problematic conviction that another individual is the subject of certain mental states is not inferred deductively from his behaviour.... Rather that conviction is a singular explanatory hypothesis of a perfectly straightforward kind. Its function, in conjunction with the background laws of folk psychology [intentional explanation] is to provide explanations/predictions/understanding of the individuals' continuing behaviour, and it is credible to the degree that it is successful in regard to competing hypotheses.¹⁷

Churchland is quite explicit that intentional explanation is a theory whose terms refer to internal states of an agent. "Each of us understands others as

¹⁷ (Churchland 1981)
well as we do because we share an integrated body of lore concerning the lawlike relations holding among external circumstances, internal states, and overt behaviour."18

1.5 Dennett on Pattern Recognition

Intentional psychology is the practice of recognising typical patterns in the purposive actions of agents and attributing to agents the beliefs and desires which best explain those patterns. According to Churchland, we recognise those patterns because we have acquired a theory, the theory of intentional psychology, which makes certain features of the environment, namely patterns in the behaviour of fellow humans, salient to us. It follows from Churchland's argument (of which more in the next chapter) that if we were to acquire a new theory, different patterns might become salient to us. At this point I only wish to note that if intentional psychology is, as Churchland suggests, a theory of other minds, humans must acquire and use that theory quite early in their development, because it seems that even very young children are competent with the concepts of intentional psychology. Developmental psychologists who support the "theory theory" have argued that young children's cognitive development shows stages typical of theory acquisition in general and that their behaviour patterns are evidence of the progressive adoption of more sophisticated theories of other minds. It is this interpretation of evidence I wish to dispute. However, before turning to the specific case of early childhood, I need to show how it might be possible that someone could acquire and deploy a set of intentional concepts without thereby acquiring a theory in the sense advocated by Churchland.

Interestingly, although he mounts a cogent argument for treating intentional psychology as a Folk theory which defines mental states as the causally efficacious brain states which cause behaviour, Churchland is not a

18(Churchland 1981p71)
functionalist. He thinks intentional psychology is, in fact, a false theory, whose flaws are disguised by its superficial structural mimicry of the more scientifically respectable theories to which he compares it. He predicts that intentional psychology will turn out, like Phlogiston theory, to have no subject matter: that our best descriptions of the actual brain processes and mental representations which cause behaviour will not map smoothly, or even jaggedly, onto the predicates of propositional attitude psychology. So the semantics of folk psychology are misleading. Because its key terms are defined in the same way as theories in other domains, we expect it to have the same predictive and explanatory power. However, because the theory has no metaphysical substance it is, according to Churchland, explanatorily vacuous.

Dennett is someone who agrees that it is unlikely that intentional psychology "limns the true and ultimate" structure of the brain states which cause behaviour. However, he still maintains that it can function as an explanatorily useful, perhaps essential, set of concepts for interpreting and predicting behaviour.\(^\text{19}\)

His example is the game of Life, played on a computer screen which illuminates pixels according to very simple rules iterated \textit{ad infinitum}. As a result, determinate patterns are formed. Miniscule variations in the starting conditions, the number and pattern of pixels initially illuminated, or the rules of iteration, make massive differences in the patterns, or succession of patterns, discernible on the screen. For example, by changing one pixel from off to on in the initial configuration, one might produce, after thousands of iterations, an elephant shape instead of a Wittgensteinian Duck-Rabbit. These patterns, Elephant or Duck-Rabbit, are discernible for humans with a bit of zoological knowledge, but from another perspective may just appear as meaningless blobs on the screen. Nonetheless, Dennett says, they are real patterns.

At this point Dennett seems to be ignoring a distinction between the reality of a pattern and its realisation. The pattern (being shaped like an

\(^{19}\)(Dennett 1991)
elephant) is an abstract object. That same abstract object can be realised by a piece of knitting or by a pile of oranges, by any collection of objects organised in the right way. The pattern specifies how the objects must be organised if they are to produce an elephant shape. With this qualification in place we can agree, not that the pattern is real, but that the pattern is realised whenever it is physically instantiated, in this case by illuminated pixels.20

Dennett makes another point, using the example of the game of Life, about the relation between patterns and the processes which produce them. The outcome of the process is a considerable degree of complexity and organisation, discernible as patterns. The perspective from which patterns are recognised, a certain cognitive and perceptual orientation, may not be the best perspective from which to derive the algorithm which produces the patterns. The algorithm, which produces the succession of patterns on the screen, by defining a function which maps from one sequence of pixel illumination to the next, is not itself presented to the pattern recogniser. So it is quite conceivable that one could, after observing the sequences of patterns over a long period, become an excellent predictor of the sequences without thereby gaining any insight into the processes which cause the sequence of patterns.

The same may be true of intentional psychology. The recognition and interpretation of behavioural regularities as patterns of purposive action requires us to treat each other as intentional agents, possessed of desires to achieve certain goals and beliefs about the best way to achieve them, given the context in which we have to act. As the example of the game of Life shows, the causal processes which bring about the patterns to which our macro level recognitional capacities are attuned, need not be structurally isomorphic to the patterns discernible at the intentional level. The game of Life showed that we may be sensitive to patterns, even be able to predict them, without thereby purporting to describe the causal processes which underlie them. The moral for

20Of course what I say here crudely truncates discussion of ontological issues raised by the use of abstracta to predict and explain the physical environment, which Dennett explores in much more detail in 'Real Patterns'. For a reply to Dennett on the ontological question see (Haugland 1993; Haugland forthcoming)
intentional psychology which we should draw from Dennett’s example is that the causal processes, described by neuroscience or modelled by cognitive science, which bring about the patterns we describe using intentional vocabulary may be similarly inscrutable at the intentional level. So for the patterns of intentional psychology to be realised does not require that the ultimate causal processes which realise beliefs and desires of agents be isomorphic in structure to the pattern discernible at the intentional level.

The neo Quinean claim was, however, that intentional explanations are rough and ready theories about the nature of the causal processes which realise the patterns which we explain using intentional idioms. One reason for adopting the neo Quinean approach is that it allows us to leave scepticism about other minds aside. Dennett’s approach to the sceptical issue in ‘Real Patterns’ is slightly different. He wants to say, I think, that once we distinguish pattern recognition and prediction, from the explanation of the causal processes which realise the patterns, the sceptical issue loses its troublesome appearance and ceases to be interesting.\(^{21}\) What is interesting is the scientific explanation of the patterns which are salient to us as rational language — using creatures with a distinct evolutionary and cultural heritage. Perhaps there is another perspective from which human behaviour, the very same set of causal processes, would present itself as a different set of patterns. If one could adopt that different perspective a radically new interpretative strategy, employing a different interpretative vocabulary, would evolve. According to Dennett there is no way to judge which of the two interpretative schemes is true.

1.6. Logical Behaviourism and the Dispositional Account of Mental States

Dennett is heir to a philosophical approach, that of Gilbert Ryle, which similarly seeks to defuse the possibility of scepticism about other minds while

\(^{21}\) (Stich, 1992), seems to think that, once we distinguish the pattern-recognising type of explanations from those which deal with causal processing, the whole issue of their relation becomes uninteresting, and the project of explaining the patterns in terms of their causes misconceived.
vindicating intentional psychology, without regarding it as a substantive
type of reasoning about the causes of behaviour. Ryle’s analysis of intentional explanation
is very similar to Dennett’s in two ways. Firstly, it regards intentional explanations as systematisations of the behavioural dispositions of agents and, secondly, it abstracts away from the causal processes which actualise or realise the patterns of dispositions.

Logical behaviourism is the view that mental states are dispositions of an
agent to behave in certain ways. Dispositions are capacities, propensities or
tendencies which are actualised in certain conditions. Solubility is a classic instance of a dispositional property, as is fragility. Things which are soluble dissolve under certain conditions and things which are fragile tend to break under certain conditions. Logical behaviourism relies on a fact about intentional explanations: to explain someone’s actions in terms of her mental states is to attribute to her certain dispositions. Someone who believes that $p$ is disposed to infer $q$, $r$ and $s$ where those are the consequences of $p$. To desire that $p$ is to be disposed to act so as to bring about a state of affairs specified by $p$. As Ryle said:

to believe that the ice is thin is to be unhesitant in telling oneself and others that it is thin, in acquiescing in other people’s assertions to that effect, in objecting to statements to the contrary, in drawing consequences from the original proposition and so forth. But it is also to be prone to skate warily, to shudder, to dwell in imagination on possible disasters, and to warn other skaters. It is not only a propensity to make certain theoretical moves, but to make certain executive and imaginative moves, as well as to have certain feelings.

22 Of course not everyone believes all the consequences of her beliefs. For example few people are disposed to assert the the Banach Tarski theorem even though it is a consequence of some basic geometrical beliefs. Intentional explanations recognise this and require only that people be treated as rational to the extent that they are purposive agents. We tolerate considerable irrationality, when measured against the canons of logic, from our colleagues but when their behaviour becomes inexplicable on the assumption that they are acting to satisfy their desires according to their beliefs they are irrational, not just theoretically but practically. This is why Davidson says “in the case of incontinence [failure of practical reason] the attempt to read reason into behaviour is necessarily subject to a degree of frustration.”

23 (Ryle 1949 p135)
Ryle was right to say that the practice of intentional explanation depends on the ability of human beings to accurately describe each other's tendencies, liabilities and propensities in actual and counterfactual situations. A belief that \( p \) rationalises all the theoretical, executive and imaginative moves, evidenced in behaviour, which follow from the content of \( p \). As Ryle put it, they all hang on a single propositional hook. So to say that \( S \) believes that \( p \) is to say that, *ceteris paribus*, \( S \) is disposed to behave in ways which are rational according to the content of \( p \) and the contents of \( S \)'s other beliefs and desires.

Ryle's logical behaviourism is an ontological thesis: that mental states are dispositions, not the mental or bodily states which cause the behaviour which counts as the actualisation of the disposition. It is also a semantic thesis: that mental state terms refer to dispositions, not the causes of the behaviours which count as the actualisation of the disposition.

At this point it is worth making a distinction between two ways of understanding the notion of a disposition, emphasised by D.M. Armstrong. The first way is to treat dispositions as "bare" dispositions, that is as tendencies which may be attributed to objects, things or persons irrespective of their intrinsic features. On this view two things which are intrinsically identical may still differ in their dispositional properties. Identical twins may differ in that one has the disposition to be called Irvine by his family and his brother has the different disposition to be called Michael. In this sense of disposition, one might have many propensities which do not depend on one's intrinsic nature and one can acquire or lose them without any alteration to one's intrinsic properties.

Now consider dispositions like solubility or electrical conductivity. Provided the background conditions do not change (conditions like the laws of nature, temperature and gravity remain constant) objects cannot acquire or discard dispositions like these unless there is some alteration in their intrinsic makeup. For example superconductors of electricity lose their conductivity...
when outside a certain temperature range because of changes to their intrinsic physical structure which occur unless they are kept at constant temperature. Another example is the disposition of water to freeze at 0 degrees, a disposition which it bears in virtue of its physical structure and which it cannot gain or lose in the same way it can gain or lose the property of being the medium in which baptisms are conducted.

Sometimes dispositions are discussed as if they are the "bare" sort and other times treated as properties objects bear in virtue of their intrinsic structure, such as the tendencies of objects to conduct electricity, to melt, or perhaps to change in characteristic ways in response to environmental stimuli. It is not always clear which of these usages Ryle is employing in discussions of, for example, belief, although he is often treated as identifying states of mind with bare dispositions.

The distinction between bare and categorical dispositions allows us to redraw, in dispositional terms, the distinction between the semantics and metaphysics of mental state terms, raised by consideration of Churchland's argument. In fact, except for "Cambridge" dispositions like those cited in the initial characterisation of bare dispositions, one will have all one's dispositions in virtue of one's intrinsic features. So one will be disposed to behave in certain ways because of states of one's brain and body. Ryle did not dispute this but he claimed that mental state terms referred to the consequent dispositions, not the states of the categorical base. This is the semantic issue. The metaphysical issue, as Churchland correctly put it, is determining the nature of mental states, which entails that a semantic theory encoded in everyday practice might be substantially incorrect. Of course Ryle thought that the semantics and metaphysics were consistent: mental state terms refer, correctly, to dispositions. Armstrong thought they referred to categorical bases. Churchland thinks they "purportedly" refer to states of the brain with a determinate sentential structure which are cited as the antecedents in psychophysical laws.
To make the distinction clearer, consider a case uncomplicated by mental language. Imagine a piece of metal with only three dispositions: to conduct electricity, to melt at 400 degrees C and to shatter at -400 degrees C. It has each of these dispositions in virtue of its microstructure. Each of the actualizations of a disposition is the outcome of the instantiation of a causal law, which mentions as antecedent the microstructural state of the metal. Someone who knew the intrinsic nature of the metal, plus the relevant laws, could predict all the metal’s dispositional properties. If the disposition were conductivity then its actualisation is an instance of a general law relating microstructure (M) to conductivity (C):

\[ \forall x (Mx \rightarrow Cx) \]

The melting, conducting or shattering behaviour can thus be explained as an instance of the application of a law which mentions the microstructural properties as antecedent. Conducting electricity comes out like this:

\[ \forall x (Mx \rightarrow Cx) \land Mx. \]

Each other disposition of the metal can be explained in the same way. In each case the antecedent in the causal law will be the same, the microstructural property of the metal, but the consequents will differ.

That microstructure is intrinsic to the metal. It can be described without reference to the actual or counterfactual environment. That description, plus the relevant laws, enable us to establish all the dispositions the metal has. In such cases reductive ambitions are encouraged. A scientist might do well to regard the bare dispositions, conductivity, melting, and fragility as evidence for an underlying intrinsic state of the metal which causes the dispositions to be actualised according to the relevant laws. Such a scientist might say that to explain the behaviour of the metal is not to systematise its bare dispositional properties but to use them as evidence for the nature of the microstructure of the metal.
Clearly, conductivity, melting and fragility are dispositions which can be analysed in more than one way. Bare dispositions can be attributed without making any commitment as to the intrinsic nature of an object. Equally, however, we might understand conductivity, fragility and melting as dispositions borne by the metal in virtue of its microstructure.

Bearing in mind these facts about the relation of dispositions to causal explanations, how should we explain the behaviour of our piece of metal? Suppose that there is a predicate, D, which applies to all and only those things which possess the conductive, melting and fragility dispositions of the piece of metal. Although D is attributed on the basis of the dispositional profile of the piece of metal, the way it functions in the explanation of metallic behaviour seems ambiguous. One way to explicate the ambiguity is to ask "to what does D refer?" or "what is the truthmaker for D statements?" There seem to be at least three candidates:

**Firstly**, the tendencies, liabilities or propensities on which attributions of D are based. So the referent of D is just the bare disposition to conduct, melt and shatter realised by all concrete instances of conducting, melting and shattering which make it up.

**Secondly**, the microstructure, or categorical base, which enables the dispositions to be actualised. For example, to have influenza is certainly to be disposed to do certain things; to have headaches, painful joints, to perspire and be unable to sleep. Influenza is actually a state of infection by a virus. Consequently the bare dispositions of an influenza sufferer function as evidence for the underlying state of infection which explains the dispositional profile of someone with influenza. What we refer to, via terms we apply to the sufferer on the basis of the dispositions she exhibits, is really their underlying categorical base: a state of viral infection.

**Thirdly** we might regard D, not as an intrinsic microstructural property, or as a set of bare dispositions but as a second order property which things bear in virtue of their first order properties. On this account, to refer to the
conductivity of a metal is not to refer to its microstructure, or to its behavioural propensities as actualised, but to its ability to play a typical causal role in the transmission of electricity. The ability to play this causal role in turn depends on other properties of the metal: those which ground its causal potentialities. The causal transaction which actualises the second order property is, *ipso facto*, the actualisation of a disposition, but the predicate D refers to the ability of the piece of metal to play various causal roles, not the playing itself.

This type of explanation distinguishes the functional role (conductivity) from the first order state (the microstructure) which realises it and regards D as referring to the ability of the microstructure to play that role. If you fed a dog some strychnine, a rat some arsenic and a philosopher some hemlock they would all be in a state of being poisoned. That state is a second order property, realised in each case by different first order properties: arsenic in the bloodstream in the case of the rat and hemlock in the bloodstream in the case of Socrates. Poisoning is thus a functional role which can be played by different first order states.

For a functionalist, the predicate D, attributed on the basis of the ability of the metal to play the various roles of conducting, melting and shattering, refers to the causal properties of the metal which bring about the shattering, melting and conducting. For the functionalist, as for Ryle, "outward physical behaviour and tendencies to behave do in some way enter into our ordinary concept of mind. Whatever theory of mind is true, it has a debt to pay, and a peace to be made, with behaviourism."24

There are actually two types of functionalist, classifiable according to the way in which they derive the identity of mental states with brain states which realise second order properties. Churchland’s argument is with psychofunctionalists who argue that intentional psychology is made true by lawlike relations between states of the brain and behaviour. These functionalists treat behaviour, the actualisation of dispositions, as the outcome

24(Armstrong 1968 p68)
of causal laws linking states of the categorical base to behaviour. Thus they
treat the explanation of behaviour by intentional psychology as a matter of
inferring the nature of the categorical base by the application of the appropriate
covering law.

In maintaining this position one needs to give an account of causal laws.
Churchland would reject the account of laws on which the psychofunctionalist
relies because the extensions of mentalistic predicates are so disjunctive.
Considered in extension, the occupants of the role of belief, for example, have
no intrinsic causal powers in common. If a computer, a dog and I all believe that
the sun is shining, as a functionalist must allow it is in virtue of states of our
brains which represent this proposition. But the causal powers of a computer's
microcircuitry, a dog's brain and my brain have nothing in common.

One reply the functionalist can make is to adopt the idea that all laws,
including those of physics, quantify over functional equivalences. The idea can
be presented using Fodor's explanation of the way the term 'money' functions
in the "laws" of economics. The extension of the term 'money' is a large
disjunct, from dollars to dirhams to cowrie shells. Nonetheless it is a "law" of
economics that bad money drives out good. In fact, laws of economics are just
like laws of physics in that they are stated as arithmetical relations between
abstract objects, just as physics deals with the numerical relations between
abstracta like mass and charge. Neither type of explanation need worry about
the details of physical instantiation as long as the theory succeeds. Thus,
according to Mellor, the laws of physics, like those of other basic sciences like
chemistry, quantify over variables which have as their value certain functional
equivalences or theoretical roles. Thus in every theory which states laws one
will encounter predicates whose extension is disjunctive. The type of
disjunction which is massive and obvious at the level of economics and
intentional explanation exists in more subtle forms in all explanatory
enterprises, including physics.

25See, for example Properties and Predicates in (Mellor 1991)
Chapter 1  Understanding Other Minds

This view of physics, that its laws quantify over functional equivalences, is called by Blackburn thesis T:

Physical thinking is essentially a question of finding the one state which covers many realisations. It is forever a question of finding a unifying feature, a pattern in the evolution of systems. If this is finding a role state that permits different realisations "all the way down" there is no harm in that.26

He contrasts thesis T with the view, held by David Lewis, that physical theory because it describes the ultimate structure of reality, which is uniform, predicates properties of uniform extension:

We may think, for instance, that mass and charge are among the fundamental properties; and whatever fundamental properties remain as yet undiscovered are likewise instantiated by very small things that come in very large classes of exact duplicates.27

Of course Lewis might be right about physics and Blackburn right about all the explanatory enterprises which deal with reality at a higher level. Intentional explanation would be just one of these enterprises, relying on ceteris paribus laws, which relate predicates which can be multiply realised.

However there are other functionalists who reach the same conclusion without the invocation of psychological laws. These functionalists are descendants of Ryle, relying on conceptual analysis to show, not that mentalistic language refers to dispositions, but that it refers to the occupants of functional roles. 28 Thus, like Ryle, they maintain that mentalistic language does not state a theory about lawlike relations between states of mind which

26(Blackburn 1990)
27(Lewis 1994)
28 The classic statement of the Ramsey sentence approach to establishing the identity of brain states with mental states is (Lewis 1970). For a further discussion which exploits the device of two dimensional modal logic see (Jackson 1993) Jackson places his version of conceptual analysis in the context of recent discussions of the necessary a posteriori and a prioricity. He thus defends functionalism from one version of Churchland’s argument based on treating mental state terms as rigidly designating the structure of the brain which causes behaviour.
represent propositions and the behaviour rationalised by intentional
psychology.

Before elaborating Ryle's argument that intentional psychology is not a
set of causal hypotheses but a way of describing the interrelations among an
agents dispositions, I need to make one further distinction which separates Ryle
from both psychofunctionalists and commonsense functionalists. Both the
psychofunctionalist and the commonsense functionalist think that states of the
agent's categorical base (i.e. brain or body ) cause behaviour which counts as
the actualisation of a disposition. The former thinks that intentional psychology
states laws which describe this causal relation.

Clearly then they differ with Ryle in thinking that the practice of
intentional psychology commits us to identifying mental states in terms of
their causal powers. However in arguing for this identification they also rely on
a conception of behaviour, the *explanandum* of intentional psychology which is
different from the one used by Ryle.

Ryle describes behaviour in a loose and intuitive way as "a stretch of
intelligent activity" to be distinguished from actions which are unintentional,
like tripping, or, like breathing, automatic. He regards behaviour as essentially
a sequence of activity, which is interpreted and explained by intentional
psychology as the activation of a set of dispositions.

The functionalist, whether psycho or commonsense, however, has a
different and more precise account of behaviour in mind. He treats behaviour as
the output of a system, like numbers flashing on the screen of a calculator. He
treats perception as the input to that system and mental states as those states of
the system which mediate between input and output. To put the point in terms
of the calculator example, mental states are the causal transitions among the
calculator's intrinsic states which link inputs (like the entering of information
via the keyboard) to outputs. In the human case mental states are the states of
the organism which link behaviour, conceived of as output, to perceptual input.
The distinction between conceiving of behaviour as output and
dehaviour as a sequence of activity is important in two respects. The first is that
conceiving of behaviour in this way is congenial to an analysis of intentional
psychology as a set of causal hypotheses about the origins of outputs. Ryle's
response to this analysis will be discussed in the next section. The second is in
the constraints the output conception puts on theories about the representative
properties of the intrinsic structures which cause behaviour. More on this in the
next chapter.

1.7 Ryle's Attack on Theory

On Ryle's view, the rationalising role of intentional psychology requires
only that the patterns of intentional behaviour given by the inferential relations
of the propositions which describe the relations among an agent's various
dispositions, be physically realised, without endorsing any hypothesis as to the
way in which that realisation comes about.

Ryle's book is called the Concept of Mind. He is trying to establish, by
close analysis of the practice and structure of intentional explanations, what
type of concepts they employ, or are committed to, in making sense of
behaviour. In particular, he wants to show that intentional explanation is not
committed to any Cartesian account of the mental. There are two aspects of
Cartesianism he attacks directly by his identification of mental states with the
dispositions of an agent. Firstly he undermines the temptation to dualism in
explaining the mental. The mental life of human agents is explained in
dispositional terms as patterns of the interpretable public behaviour of
corporeal human agents. Nothing about this invites the postulation of a
mysterious inner realm, accessible only in introspection, ontologically distinct
from the bodies and behaviours it is supposed to explain. On the dispositional
account the mind is not hidden behind the behaviour of other people but is
directly accessible in the patterns of behaviour themselves.
Secondly, he attacks both the Cartesian and the related, neo Quinean, view, that the mind, while it may not be a separate substance, is, nevertheless, something which lies behind the patterns of overt behaviour through which it is exhibited: states of the brain which cause the behaviour. Once again the dispositional analysis is employed to show that intentional explanation is not a matter of inference from observed behaviour to a different state which lies behind, and causes the behaviour. Intentional explanation is exhausted once the logical relations among dispositions have been correctly characterised. Intentional states "hang on a common propositional hook" because of the actual and counterfactual patterns of behaviour which they rationalise. The content of the state specifies the dispositions which are actualised in behaviour, by fitting behaviour to a pattern. The dispositional analysis is thus fatal to any view of mental states which regards them as intrinsic physical states of the brain or body which cause behaviour. Since a mind is just the dispositions of an agent related to each other according to their contents, no term used in intentional psychology refers to an inner state, material or immaterial.

For the same reason, he attacks the identification of mental states with causes of actions. Causes of actions must be the brain or body states which cause behaviour, but such states are inner, intrinsic states which lie behind behaviour, they are certainly not identical with the tendencies of an agent to behave, even though those tendencies depend for their actualisation on causal processes.

It should be clear that Ryle is not proposing a reductive behaviourism. He does not think that mental terms can be translated without residue into behavioural terms. After all, on his view, one cannot even characterise dispositions without using mental terms, as his analysis of the belief that the ice is thin shows. That analysis mentions dispositions to imagine, form other beliefs and to be apprehensive, all of which are mental abilities. So the dispositional analysis is not reductive.²⁹ Nor is Ryle proposing materialism as

²⁹Note however that reductive analyses in terms of behaviour, or for that matter neurophysiology, can be achieved even though individual mental states cannot be
an alternative to a dualist account of the mental. He is saying that the materialist/ dualist question only arises when we commit the category mistake of thinking that mental terms, which explain behaviour, do so by pointing us towards other states which lie behind the behaviour, inner states, material or immaterial, for which behaviour is evidence.

Ryle is committed to the view that intentional explanation exhibits the rational connections among behavioural dispositions of corporeal agents. Such an explanation need make no mention of the causes of behaviour or of the relation of intentional psychology to other explanatory enterprises such as the science of brains and behaviour. This is why he can say that explanations of behaviour in terms of mental states are independent of causal explanations of behaviour.

Ryle does seem to allow that some mental states, "feelings" and "dwelling in imagination" can be inner and are recognised as such in intentional explanation. But this is consistent with his overall picture which is based on conceptual analysis. Obviously tendencies to behave in certain ways are not an essential part of our everyday concept of imaginings and feelings. Our concepts of belief and desire, which include the concept of direction of fit, invoked in the explanation of action, are based on the tendencies of agents to behave so as to satisfy the content of the belief or the desire. The emphasis on propositions in intentional explanation thus derives from the interpretation of behaviour as rationally directed toward solving problems set by the environment. There is thus no need to regard propositions as attempts to capture the nature of the state which causes behaviour. Propositions specify the logical link between dispositions, not the causes of the behaviour which counts as actualising the disposition.

---

characterised independently of mental concepts. Reduction via Ramsey sentences achieves this result. This is, in fact, the project outlined by Lewis in 'Psychophysical and Theoretical Identifications'. The idea is briefly that one can write out as a, possibly infinite, disjunction all the mental truths, characterised in predicates of intentional psychology and then show that they are entailed by all physical truths, expressed as predicates of neurophysiology or physics. The project thus explains the mental in terms of physical facts without requiring that each truth of psychology be independently characterised as a truth of physics or neurophysiology. See note 27.
Note however that this does seem to show that, whatever Ryle's intentions, the dispositional conception is best thought of as a categorical one, rather than the identification of mental states with a set of "bare" dispositions. Ryle wants to explain the actual and possible behaviours of corporeal agents. Those behaviours must be the behaviours of the agents themselves, thus they must depend on states of the agent's brain and body, not on the actions of other agents or states of the world extrinsic to the agent. Which of my dispositions are explained by intentional psychology? Those whose actualisation is the outcome of a causal process which originates with me. One can believe this without being committed to the view that the explanation of behaviours states hypotheses about the nature of those causes. Behavioural patterns, explained by intentional psychology, might stand in the same relation to their underlying causes as the patterns in the game of Life stand to their underlying causes. In such a situation we would not regard the practice of pattern recognition and prediction as entailing any hypothesis about the nature of the processes which produced it.

One might have sound empirical reasons for believing this: one might doubt whether there could be states of a brain which could causally explain the rational connections among all the actual and counterfactual behaviours of a human agent. But Ryle's arguments are designed to show even if a reductive account of belief or desire in terms of causal processes of brains and bodies could be provided, it would still be a mistake to regard intentional explanations as referring to those properties and processes. "Overt intelligent performances are not clues to the workings of minds, they are those workings." 30

Rylean, and Dennettian, intentional psychology require only that we be sensitive to patterns or connections among typical agent behaviours and express those patterns using propositions whose inferential connections parallel the relations between the agent's dispositions. We are not using propositions to report facts about the structure of the agent's brain, but facts

30 (Ryle 1949 p58)
about the agent's behaviour in actual and counterfactual environments, which can be ascertained independently of any facts intrinsic to her brain or body.

Although the dispositional account addresses the problem of other minds by making mental operations public and accessible, there are other aspects of intentional psychology which fit more readily with a conception of minds as the site of internal states which lie behind behaviour. The first point to note is that intentional explanation, as Ryle himself noted, deals with "stretches of intelligent behaviour." This is part of what is meant by the "rationalising role" propositions play in intentional explanation. An automaton which mimics my behaviour under remote control is not rational because it is not making any decisions but just being caused to make certain movements. We sometimes describe this feature of rationality as a capacity for deliberation. An agent who believes that $p$, desires that $q$ and does $r$ as a result is not exhibiting stretches of character and intelligence unless she is able to realise that doing $r$ is rationalised by the content of $p$ and $q$.

But what is the site of such a realisation if not the mind of the agent? And, apparently, this mind must be considered as something distinct from the behaviour which it monitors and whose outcomes it ponders. The requirement of rationality seems to drive the conception of mind back inside, no longer something merely manifest in behaviour but lying behind that behaviour. This is especially so when we consider that the obvious candidates for mental states which are involved in such acts of realisation and deliberation are imagination, speech in foro interno, or some other type of information processing which allows the contents of mental states to be accessible to the thinking (though not necessarily conscious) mind. (These considerations of rationality will play a part in later chapters. In chapter two we will consider Ryle's arguments that rationality can be explained without postulating any "inner" activities. In chapter three an account of the acquisition of psychological concepts is shown to depend the assumption that intelligent behaviour is to be explained by the postulation of intentions conceived of as inner states. In particular, certain
theories of developmental psychology explain autism as an inability to understand that others have an "inner" mental life which causally explains their behaviour. Chapters four and five reevaluate this conception of the autistic disability, reinterpreting the evidence on which it relies, using a conception of the nature of intentional concepts which develops from the Rylean one outlined here.)

The second objection to a purely dispositional account can be stated quite simply. Recall the conclusion we drew from Dennett's game of Life, that the phenomenon of pattern recognition places no constraints on the causal processes which bring the pattern about. But intentional explanation must place some constraint on the causal processing which realises the dispositions which it tracks via its attribution of mental states. Exactly what these constraints are is the issue in philosophy of mind. Nonetheless we can be quite sure that there are some, because as those descendants of Ryle, the Commonsense Functionalists, point out, not just any physical system whose causal inputs, internal transitions and outputs sustain behaviour identical to that of an intelligent human being, will qualify as an intentional system. A computer programmed with all the information needed, and typical behaviours produced, by a human at each stage of its life could produce behaviour indistinguishable from that of its human counterpart. The computer could just "look up" its information store and select an appropriate behaviour for each situation. However, all sides of the artificial intelligence debate are agreed that human intelligence and rationality could not be realised by such a system. For the same reason, we would not regard someone who could only produce answers to written mathematical questions by punching the symbols into a calculator and reading off the result, as exhibiting an ability to reason out mathematical problems. There seems to be consensus that, for similar reasons, rationality is not realised by the purely formal manipulation of symbols according to rules.

31 Such a computer is, of course, the infamous Blockhead. (Block 1981)
Dennett is very hard to pin down on this issue. In some places he says that two systems which are behaviourally identical are psychologically identical. ("let me confirm Jackson’s surmise that I am his behaviourist, if two organisms are exactly behaviourally alike they are psychologically exactly alike.")32 If this is the case then Dennett needs to say why it is that those who think Blockhead and a human who produce identical behaviour are not psychological duplicates are wrong. Jackson and his fellow commonsense functionalists have an answer: Blockhead has no internal functional architecture. One can characterise neither his attitudes, nor the content of his states, in terms of their functional profile. Dennett’s characteristic response here is to bite the bullet. What is so wrong with saying that Blockhead has belief? If saying so violates a few intuitions that is a small price to pay for the increased understanding of mental phenomena in general. Note, however, that there is almost universal agreement that Blockhead is not a fit subject for intentional explanations.

Ryle has clearly drawn a conceptual distinction between dispositions and internal states which play a causal role in the production of behaviour. However, as I said above, drawing that distinction cannot tell us whether intentional explanation deals in dispositions alone, or whether it treats dispositions as evidence for internal states which cause behaviour. Ryle saw the issue (prior to the advent of functionalism and the identity theory of mind) as a matter of determining which of the dispositional theory and the internal state theory, as it figured in Cartesian accounts of mind, best fitted our concept of mind. Ryle saw the immaterial aspect of Cartesianism as a by product of the application of standard scientific method. As he says:

a person’s thinking, feeling, and purposive doing cannot be described in the idioms of physics, chemistry and physiology, therefore they must be described in counterpart idioms ... as the human body, like any parcel of matter, is a field of causes and effects, so the mind must be

32(Dennett 1993p923)
another field of causes and effects, though not (heaven be praised) mechanical causes and effects. 33

Ryle was criticising any view of mind which treated behavioural explanation as inference from observed behavioural effect to unobserved behavioural cause in a state of the agent's mind. He says: "if there were such inner states and operations one person would not be able to make probable inferences to their occurrence in the inner life of another." 34 One reason why Ryle would resist this analysis of mental states is the dispositional account: intelligent behaviour is not evidence in support of a hypothesis about a mind with certain properties — it is the mind itself in action. But that cannot be his reason here. Here he is saying that, even if behaviour is caused by states of mind which represent propositions, we could not understand other people by inferring from their behaviour to the content of their mental states.

Understanding a person's deeds and words is not, therefore, any kind of problematic divination of occult processes. For this divination does not and cannot occur, whereas understanding does occur. 35

Ryle is saying that intentional explanation does not treat dispositions as evidence for inner states. He constantly expresses this by a medical metaphor: the dispositions of an agent do not stand to mental states in the same relation as symptoms stand to states of infection: intentional explanation is not the equivalent of diagnosis.

One reason Ryle gives is that diagnosis is a scientific procedure which depends on the existence of laws connecting the states of infection to the symptoms which a doctor observes. A doctor is justified in inferring from the evidence presented via certain symptoms to the cause of those symptoms.

33 (Ryle 1949p18)
34 (Ryle 1949p54)
35 (Ryle 1949p55)
because the symptoms are connected to a state of infection by well confirmed biochemical laws. According to Ryle, however, there are no laws of psychology, no laws which connect states of mind with behaviour, and the successful practice of intentional explanation proceeds regardless. Intentional explanation cannot, therefore, be based on the exercise of anything like an inference, licensed by the laws linking behaviour to states of mind, from behaviour to states of mind. He thus denies the first premise in Churchland's argument which states that intentional psychology exploits a covering law model of scientific explanation.

However, even if there are no psychological laws, their absence does not establish the thesis that intentional explanation does not purport to state laws. It may well be the case that intentional explanation is constructed and conducted as if there were psychological laws, despite their absence. This is the position Churchland maintains: that intentional psychology is a bogus theory.

It might seem odd to place Churchland and Ryle together as theorists of the mind, but they share a common preoccupation with the everyday conception of mind and its relation to the causal explanation of behaviour. Both acknowledge that behaviour, the raw material of intentional explanation, is caused by inner states of the agent. They both face the following problem. How can we know what other people are thinking, which propositional attitudes do they have, on the basis of their behaviour? Ryle argues that there are no causal laws linking minds and behaviour. Mental states terms cannot, therefore, refer to the causes of behaviour. Analysis of the concepts used in intentional psychology reveals that mental states are dispositions. Churchland argues that, since behaviour is caused by inner states, intentional explanation relies on a network of laws about the relation of those inner states to behaviour. On either view, there is no essential mystery about the nature of mind created by the practice of intentional explanation. Both think that intentionality can be explained, in purely material terms, once the role of propositions in intentional explanation is accounted for. In this respect they share a hostility to
introspection as a reliable guide to the content of thought, and as any part of
the explanation of intentionality, because of the sceptical problem it creates. As
a logical behaviourist, Ryle regards propositions as expressing the inferential
links between terms which refer to dispositions. Intentional explanations are
thus a species of analytic inference, not of hypothetico-deductive reasoning.
Churchland, on the other hand, treats inferential relations among propositions
as functioning in intentional theory the same way as numerical relations among
theoretically postulated abstracta like 'mass' or 'velocity' function in any
scientific theory.

1.9 Structure and Theory

There is no doubt that humans acquire and store a vast amount of information
which they can retrieve and use to think systematically and then express these
thoughts in language. So the brain has some structural architecture, because
this type of order does not proceed out of chaos. (Though see Wittgenstein
below) At the very least we are able to show that some mental activities
typically depend on brain activity, which shows that there is at least a minimal
level of functional organisation associated with thought.

Related to this point is another, crucial to the explanation of
intentionality. The brain stores and retrieves information about the external
world and about states of the organism in which it functions. Therefore it can
be said to represent propositions which are made true or false by states of the
world. If those representations were in sentential or pictorial form and
processed according to rules of natural language then there would be an
obvious candidate for the truthmaker for intentional explanation in the neural
structure of the brain. In this case the brain would represent information by
neurally encoding propositions. Those who reject this account but wish to
defend intentional psychology need to explain if any, degree of organisation
and structure of information processing in the brain would count as the
physical realisation of intentionality.
Wittgenstein sometimes appeared to argue that explaining behaviour in terms of structures and processes within the brain was some sort of category mistake, derived from a misguided attempt to understand intentional explanation as a science of the mind:

Thinking in physiological hypotheses deludes us sometimes with false difficulties, sometimes with false solutions. The best prophylactic against this is the thought that I don’t actually know whether the humans I am acquainted with actually have a nervous system. (Remarks on the Philosophy of Psychology. 1.106)

This seems to be compatible with the bare dispositional account since it implies that there is nothing about the brain or body of the agent we might discover which could shed light on her mental states. Similarly,

Strange coincidence, that every man whose skull had been opened had a brain! (On Certainty 207)

These quotations give the impression that Wittgenstein thinks intentional action arises causelessly, but perhaps this is not the case.

No supposition seems to me more natural than that there is no process in the brain correlated or associated with thinking: so that it would be impossible to read off thought processes from brain processes. I mean this: If I talk or write there is, I assume, a system of impulses going out from my brain and correlated with my spoken or written thoughts. But why should the system continue further in the direction of the centre. Why should this order [patterns of purposive activity and the employment of language with a logical structure] not proceed, so to speak, out of chaos?

The use of the expression “out of chaos” adds weight to the interpretation of Wittgenstein as someone who regarded brain structure as irrelevant to the
explanation of behaviour. Others, however, have treated his remarks as less radical, denying only that intentional activity derives its meaning through similarity of structure to the brain states which cause it. On this interpretation, his remarks about the irrelevance of physiology are designed to relocate meaning away from structure in the brain into patterns of activity, and the kind of responses to those patterns we learn when we participate in a shared culture. The purpose of intentional explanation is to articulate these responses, not to frame hypotheses about the structure of the causal processing which underlies them.

Like Wittgenstein, Ryle regards intentional explanation as a non-theoretical practice of behavioural interpretation which abstracts away from the details of causal explanation. He inaugurates the program, since followed by all reductive physicalists, of explaining intentionality and its connections with rationality and deliberation, by accounting for the use of propositions in intentional explanation. Dennett follows Ryle in this regard. He regards intentional explanation as made true by behavioural patterns realised by causal processes at a lower level. Dennett is, however, not quite as easy to interpret as Ryle on the role of propositions in intentional explanation. He has described himself as an instrumentalist, but this is misleading because instrumentalists are normally people who believe that theoretical entities are really just useful fictions or abstracta invoked to help in calculation (as an electrician talks about imaginary numbers when predicting the flow of electrical current.) While Dennett does agree that propositions are things we invoke to understand, predict and explain each other’s behaviour, he does not regard propositional attitudes as fictions, nor, he says, does he regard intentional explanation as a theory but as an interpretative practice conducted from the intentional stance. His latest view, however, seems to come very close to Churchland’s in regarding mental states with content as erroneous theoretical posits. That is to say he regards intentional explanation as a theory which uses propositions to
refer to brain states, which in fact do not have the properties attributed by the theory.

The multidimensional complexities of the underlying processes are projected through linguistic behaviour, which creates an appearance of definiteness and precision thanks to the discreteness of words.

He goes on to quote Churchland, who does regard intentional explanation as an error theory, in support. In the relevant passage Churchland says “Being projections of that inner reality, such sentences do carry significant information regarding it and are thus fit to function in a communication system.” However Dennett does not quote Churchland’s next sentence:

On the other hand, being subdimensional projections, they reflect but a narrow part of the reality projected. They are, therefore, unfit to represent the deeper reality in all its kinematically, dynamically and even normatively relevant aspects. That is to say, a system of propositional attitudes such as FP, must inevitably fail to capture what is going on here, though it may reflect just enough superficial structure to sustain an alchemylike theory among folk who lack any better theory... The real laws governing our internal activities are defined over different and more complex kinematical states and configurations... (my italics)

Dennett, however, while he endorses the last sentence, does not see intentional explanation as an error theory, because he abstracts away from the details of realisation. But one can only do this if one regards intentional explanation as made true by patterns of behaviour, rather than the structure of brain states, which would be consistent with much of Dennett’s overall oeuvre. If, however, as seems to be the case in this passage, one treats intentional explanation as a theory of brain processes, and the theory attributes those processes with a linguistic structure which is in fact not there, then the theory is incorrect, non?

36(Dennett 1991p45) Dennett’s invocation of Churchland at this point seems to support this interpretation.
If pressed, Dennett should revert to the more behaviourist construal which would allow him to abstract entirely from details of structural implementation. Alternatively, Dennett could simply identify himself as a commonsense functionalist. While this is a tenable position, one of those intermediate between the non-theory, non-structure position of Wittgenstein and positions like that of Churchland, which insist that intentional explanation states laws about brain states with propositional structure as the antecedent in causal explanation, it is not, of course, unique to Dennett.

It should be clear that where one stands on this issue depends on whether intentional psychology is a theory about the structure of the internal representations which cause behaviour. If it is such a theory, there are further questions which arise about the level of detail to which it is committed and whether the representations over which it quantifies are the antecedents in lawlike explanations of behaviour. The argument in support of the view that intentional explanation is a theory rests on three considerations: firstly the Quinean one that all terms get their meaning from their place in a network of empirical theory, secondly the argument from the isomorphism of structure between intentional and hypothetico deductive explanation and, thirdly, the argument that we can treat scepticism about other minds and the problem of intentionality simultaneously by accounting for the role of propositions in a predictive theory of rationalising explanation.

Churchland would not disagree that it functions as a set of concepts which we employ to explain behaviour, but says that if we treat those concepts as scientific hypotheses they will turn out to be explanatorily vacuous. Thus he agrees with Ryle and Dennett that there is no smooth intertheoretic translation available between the theories of folk psychology and scientific psychology but disagrees with them in claiming that it follows that intentional psychology is a useless explanatory tool.

Clearly, then, a lot turns on determining whether intentional psychology is a proto scientific theory which stands or falls on whether its predicates limn
the structure of the brain states which cause behaviour. In the next chapter I examine objections to this characterisation of intentional psychology and the alternative accounts of intentional psychology developed by Ryle, Dreyfus and Churchland *malgre lui*.

A crucial feature of these atheoretical accounts is the way they deal with the problem of rationality. I mentioned earlier that rationality was a problem for dispositional accounts like that of Ryle and Dennett because intentional psychology is the interpretation of rational behaviour. A rational person is someone who can reject false beliefs, revise or strengthen her beliefs in the light of new evidence or by considering the consistency of individual beliefs with the rest of her belief set. Normally we describe this deliberative process as the forming of second order mental states.

Part of the attraction of the "theory theory" is its account of the formation of second order mental states via a process of theory application and revision. One forms beliefs about one's own or others' beliefs or desires by considering whether or not the theory applies. In rejecting this conception, I develop an alternative way of explaining second order mental states in terms of practical knowledge. Practical knowledge is not the same as Ryle's *knowhow* because *knowhow* is not propositional. Rather, practical knowledge is knowledge of propositions which a person can be said to have in virtue of their *knowhows*. The distinction is developed in the next chapter and then applied, in chapters three and four, to the evidence from child development.
Chapter 2. Theory and Practice

Intelligent practice is not a step child of theory. (Gilbert Ryle *The Concept of Mind* p26)

Precis

2.1. Tacit Knowledge. Three ways of understanding tacit knowledge.
2.1.1. Ryle on Knowhow. Ryle’s distinction between "knowing how", the knowledge which grounds skills, and "knowing that", knowledge of the truths expressed by attributions of cognitive propositional attitudes. Perhaps the knowledge on which the practice of intentional psychology relies is of the former type?
2.1.2. The Regress Argument. Ryle’s attempt to show that knowhow cannot be explained in terms of the internalisation or internal representation of knowledge that, where knowledge that is conceived of as the tokening of symbols. The regress argument works successfully against two versions of the "Rules and Symbols" approach to cognition.
2.1.3. The Decomposition Argument. A reply to Ryle which progressively decomposes rules and symbols to elementary states of a system. However, not any system which sustains input-output relations thereby qualifies as a symbolic encoder of propositions represented by the system. The Game of Life example again.
2.1.4. The Neutrality Thesis. Recognition that a system represents certain propositions does not necessitate any view of the structure of the representational media on which the system relies.
2.2. Salience and Intentional Activity. Dreyfus’ arguments against Good Old Fashioned Artificial Intelligence. Dreyfus’ arguments take the phenomenology of lived experience at face value and construct a philosophical account faithful to that phenomenology.
2.2.1. The Link Between Perception and Action. Dreyfus relies on the immediacy, or automaticity, of the link between perception and action.
2.2.2. *Lebenswelt* or Conceptual Scheme. How to explain this automaticity? Churchland, like Dreyfus and Ryle, thinks it cannot be explained in terms of the internalisation of a set of rules and symbols which capture the propositional truths about a domain. Without endorsing any positive account of the processes involved in automaticity we can note that stating as a theory the propositions known by an expert in a domain does not commit us to any account of the cognitive mechanisms which subserve that knowledge.
2.3. Peripheral Theory and Practical Knowledge. The distinction between practical and intellectual knowledge of propositions. A distinction which refines the knowing how/knowing that distinction. The essential feature is that it shows how we can be said to know the truths expressed by propositions even where we do not intellectually represent those propositions.
2.4. Practical Knowledge. The distinction in 2.3 developed by considering the rule-following argument.
2.4.1. Aristotle on Learning to be Good. Another example, Aristotle’s account of ethical knowledge.

2.5. Seeing the Soul. Wittgenstein on other minds. Another antitheoretical account which takes phenomenology at face value and thus sets the sceptical issue aside. Introduced here because the phenomenology becomes important in later consideration of the autistic case.

2.1. A Non Theoretical Account of Intentional Explanation: Ryle on Knowhow

One ability people have is the ability to state the propositions which express the facts that they know about a particular domain, to systematise them into a theory of the domain, to test the theory against new facts, to draw conclusions, revise the theory and so on. If we want to test a person’s knowledge of a theoretical domain we ask them whether they know the facts which make up that domain, or ask them to explain how the theory of which they are the master applies to certain situations. How well they know their field is tested by their ability to produce a set of propositions which describe it and a further set of propositions explaining the relations between those propositions. It is exactly this sort of ability which Churchland thinks underwrites intentional explanations:

Each of us understands others as well as we do, because we share a tacit command of an integrated body of lore concerning the lawlike relations holding among external circumstances, internal states and overt behaviour.\(^1\)

The integrated body of lore is just those propositions which a practitioner of intentional explanation would hold true of other people and their mental states. Churchland argues that even where we do not explicitly invoke these propositions in explaining others’ behaviour we nevertheless represent them tacitly. We are like people who no longer have to do the expansion to solve a quadratic equation or a simple integral. We have internalised the relevant

\(^1\)(Churchland 1981 p69)
function through practice so that when confronted by a routine instance we can apply the function without explicitly representing it.

A real area of dispute in the analysis of intentional explanation is how to understand the nature of this internalisation, or tacit knowledge. The expression "tacit knowledge" with its contrast with explicit knowledge, suggests that the theory has been submerged below the level of consciousness, but exists, stored in the mind as a set of propositions, and is used to ground inferences about the origins of behaviour.

At this point it is worth making some preliminary distinctions. What is involved in knowing an integrated body of lore, a set of propositions?

Firstly, an agent might consciously think that "p is true." That is, she might explicitly represent that proposition via some sort of mental symbol, presumably p in her native language, and assent to it in foro interno. Although we sometimes represent our knowledge to ourselves in this way, it is infrequent. Most of the knowledge which guides our activities is not consciously represented.

Secondly, that same proposition might be explicitly represented, but in a manner inaccessible to consciousness. If this is correct then the relevant mental symbol will be a sentence, or component of a sentence, in an inner language of thought. This is the position attributed to Harman in the last chapter, that thoughts expressed in natural language expressions are transformations of subconscious, mentalese.

The antitheoretical arguments I give in this chapter are directed against these two positions, so they will work against any neo Quinean who assumes that knowledge of intentional psychology is a matter of explicitly representing it as a set of rules and symbols. However not all neo Quineans make that assumption, as the Decomposition Argument, which I will consider shortly, makes clear.2

---

2 Although Harman may avoid the antitheoretical challenge if he subscribes to the Decomposition argument.
Thirdly, the knowledge might not be represented at all, consciously or unconsciously. The agent’s responses to her environment might not depend on any internal processing of information which can be correlated with the content of a set of propositions. This is the sort of thing Wittgenstein suggested when he remarked "Strange coincidence. That every man whose head had been opened contained a brain." This third position violates an intuition that what we do and say in response to what we perceive depends on what goes on neurophysiologically, and, given what we know about the role the brain has to play in thought, is empirically untenable. The same point was made in more detail at 1.8.

Before we consider arguments against the first two positions let me point out that the purpose of this chapter is to emphasise that the everyday practice of intentional psychology is neutral about the nature of mental representation. It assumes that understanding and rationality depend on the representation of propositions, but states no hypothesis about way in which the mind accomplishes that representation. Ryle’s arguments are important because they help to establish this neutrality thesis.

2.1.1. Knowing How and Knowing That

As a start Ryle directs us to the many human abilities which do not require this type of tacit or explicit knowledge that. When someone cycles to university, avoiding bumps in the road, jumping gutters, and racing to beat the traffic across the intersection, they are showing that they know how to cycle well. But their ability to do so is not a matter of mastering a set of propositions true of good cyclists. In fact one can know all the propositions true of good cyclists without being a good cyclist. One can read cycling magazines and treatises on biomechanics and physiology until one knows everything there is to know about cycling. Perhaps one could build and animate a model of Miguel Indurain and perform tests on it to see if one’s propositional knowledge about
cycling was borne out in fact. Nonetheless, one would still not know how to cycle. Miguel Indurain knows how to cycle.

More dramatically, someone who knew some elementary physics could rapidly understand all the propositions true of a good skier descending an icy couloir. They could acquire all this knowledge from a textbook on skiing without ever seeing snow. But we would not say of such a person that they know how to ski. In fact someone who had never skied, but, after passing an examination on the principles contained in Karl Gamma's *Handbook of Skiing*, attempted to ski a steep, icy, couloir, would be seriously injured. The test of knowing how to ski is how skilfully someone skis, adapting to changes in snow conditions and the degree of slope while continuing to link their turns smoothly and safely down the hill. One cannot learn how to do this out of a book. One does so in the same way one acquires any skill. By practice, imitation, trial and error, having one's mistakes corrected and trying out the particular skill in new and progressively more difficult situations.

Ryle points out that when we describe some one as knowing how to cycle, ski, play the clarinet, dance or conduct a conversation in French we are implicitly bringing to bear criteria of evaluation for that particular activity. They can be said to do the activity in question better or worse, fluently, stylishly or clumsily, but not truly or falsely. True and false are evaluations which apply to propositions, the objects of knowledge that, and not to activities which exhibit knowhow.

Whether or not the boxer plans his manoeuvres before executing them, his cleverness at boxing is decided in the light of how he fights. If he is a Hamlet of the ring, he will be condemned as an inferior fighter, though perhaps a brilliant theorist or critic. Cleverness at fighting is exhibited in the giving and parrying of blows, not in the acceptance or rejection of propositions about blows, just as ability at reasoning is exhibited in the construction of arguments and the detection of fallacies, not in the avowal of logician's formulae. Nor does the surgeon's skill
function in his tongue uttering medical truths but only in his hands making the correct
movements. ³

Ryle is not just pointing out that knowing how to box, and knowing the theory
of boxing, (which propositions are and are not true of boxers,) are different
types of knowledge. He is also rejecting an account of skill which explains it as
the progressive storing of propositions about the particular activity which are
then tacitly "looked up" by the thinking mind to govern future activity. The
idea, rejected by Ryle, is that when the boxer drops his guard after delivering a
blow and receives a counter blow his mind tacitly stores the mentalese
instruction "keep your guard up" and in future exchanges his behaviour is
governed by that tacit instruction.

A apparent statement of this sort of explanation of knowhow is found in
Fodor:

Here is the way we tie our shoes:

there is a little man who lives in one's head. The little man keeps a library. When one acts upon
one's intention to tie one's shoes, the little man fetches down a volume entititled Tieing One's
Shoes. The volume says such things as: "Take the left free end of the shoelace in the left hand.
Cross the left free end of the shoelace over the right free end of the shoelace...etc
When the little man reads the instruction "take the left free end of the shoelace in the left hand ",
he pushes a button on a control panel. As a causal consequence of the functioning of these
mechanisms, one's left hand comes to seize the appropriate end of the shoelace. Similarly,
mutatis mutandis, for the rest of the instructions. ⁴

2.1.2. The Regress Argument

As Ryle caricatures the idea, it is as if there is one stupid or animal
person, the body, acting under instruction from another intelligent one, the

³(Ryle 1949p48)
⁴(Fodor 1968)
mind. The mind ponders the information acquired by the body in the course of its causal transactions with its environment and then constructs a suitable plan which the body then executes. Intentional explanation then becomes a matter of inferring from the behaviour of the body the nature of the plan it is executing or the rules it is following, which are represented in the mind of the agent. The idea is attractive because it allows us to explain the difference between intentional activity and mechanical movement in terms of whether or not it is governed by the, tacit or explicit, prior operation of the thinking mind.

Ryle's argument against this "intellectualist legend" is intended to apply any accounts such as Fodor's which attempt to reduce knowhow to manipulation of rules and symbols. This argument, which has many avatars, is presented by Ryle in this form:

1. Behaviour is intentional, becomes action rather than mere movement, when it is regulated by propositional knowledge stored in the agent's mind.
2. The mind, however, needs to determine which proposition or set, \( S_1 \), thereof is relevant.
3. BUT the mind cannot make this correlation automatically or mechanically because that would be an instance of non intentional behavior.
4. So the mind needs another set of propositions, \( S_2 \), which tell it which propositions to apply. \( S_2 \) is something like a rule for deriving \( S_1 \), the set of relevant propositions, from all the propositions stored in the mind.
5. BUT \( S_2 \) cannot be applied mechanically or automatically for the same reason \( S_1 \) cannot.

5Ryle attacked the "intellectualist legend" grounds which are inapplicable to the Fodorian story above. Firstly, its Cartesian emphasis on the availability to introspection of the contents of the thinking mind. Secondly, its equation of thinking with interior monologue or soliloquy, when "the sequence of your sensations and imaginings is not the sole field in which your wits and character are shown; perhaps only for lunatics is it more than a small corner of that world." Fodor, of course, assumes that the relevant instructions are not necessarily accessible to consciousness, so he is exonerated from this aspect of Cartesianism.

6 (Ryle 1949p31) "According to the legend, whenever an agent does anything intelligently, his act is preceded and steered by another internal act of considering a regulative proposition appropriate to his practical problem. But what makes him consider the one maxim which is appropriate rather than any of the thousand which are not?... Must we then say that for the hero's reflections how to act he must first reflect how to reflect how to act? The endlessness of this implied regress shows that the application of the criterion of appropriateness does not entail the occurrence of a process of considering this criterion."
6. So there must be a further set of propositions, $S_3$, which effectively function as a rule for deriving the correct rule, i.e. which tell the mind how to derive $S_2$,... and so on. As Ryle puts it:

Our intellectual planning process must inherit its title to shrewdness from yet another interior process of planning to plan, and this process could in turn be silly or shrewd. The regress is infinite and this reduces to absurdity the theory that for an operation to be intelligent it must be steered by a prior intellectual operation. 7

The regress argument depends on two considerations. The first is Salience. 8 Ryle can grant that information is stored and used by the mind in producing intentional activity. Perhaps it is stored in the way Fodor suggests, as a list of sentences which describe aspects of the world and the way to perform activities successfully. However, not everything we know about the world is salient to each of our tasks. To properly control activity then, the mind must be able to select, from its vast store of information, the salient propositions. How is this selection accomplished?

Premise three of the regress argument relies on Ryle's second consideration: the fact that this selection cannot be mechanical, as in the operation of a machine or production of the right answer to a question by rote, or it will not be intentional. The correlation must come about as a result of understanding the propositions and their salience. And, according to Ryle, there is no way to explain such understanding without vicious regress.

Ryle thinks that we are drawn into the Cartesian mistake of attempting to explain intentional activity in terms of inner mental operations by our habits of introspectively planning and rehearsing actions before we perform them. So we work out the answer to a question "in our head" before we answer, or visualise the route down a ski slope and imagine skiing it before we actually

7 (Ryle 1949p31)
8 I use this term rather than the more natural sounding relevance, because relevance has already been appropriated by Gricean theorists of communication and I rely on their technical usage of the term in later chapters.
jump off the edge. But these inner imaginings and rehearsals presuppose the abilities being imagined and rehearsed: abilities we initially acquire through our participation in the outer activity.

Imagine a musician performing a clarinet concerto. As she plays she concentrates on the difficult passages, recovers from mistakes and keeps in mind the overall structure of the piece and the clarinet part in particular. Suppose she accomplishes this by keeping the score "in her mind's eye." However the fact that the musician visualises the score in this way is no more an explanation of the ability to play the clarinet than putting the music on a stand in front of them would be. What needs to be explained is how the musician produces sounds from her instrument according to the score, whether viewed on a music stand or in the Cartesian privacy of her own imagination. That ability is explained in terms of her years of practice and performance not in terms of any internal intellectual activity.

The examples Ryle gave were of arguing, playing chess and doing logic because these activities are often accompanied by conscious deliberation in foro interno. The Cartesian temptation is to assume that the difference between intelligent and mechanical behaviour is that the first is the result of deliberation such as typically accompanies our more abstract calculating activities. And this is a natural temptation. Often when we want to determine whether someone understands an activity such as chess playing or reasoning we ask them to explain how they decided to make their move or give their answer. If they can explain their activity in terms of the application of some governing rule or principle that is prima facie evidence that their acting correctly was not a fluke. However, as Ryle points out, such an explanation will only serve to separate intentional from nonintentional behaviour if the rule or principle is understood and not merely applied by rote. Thus the intentionality of chess playing cannot be explained by saying that it occurs as the outward expression of an inner act of mental chess playing. Such an explanation leaves the intentionality of the allegedly prior inner act unexplained.
Ryle points out that we knew how to argue before Aristotle wrote down the rules of syllogistic reasoning, and almost all children learn to speak their language without explicit instruction in its grammatical rules. Furthermore, the theory of such activities, the grammar for a language, or the rules of logic, are usually acquired after the relevant ability. In fact the test for the correctness of a grammar, or a proposed logical rule, is how well it conforms to the speaking of a language or the practice of argument, things we already understand how to do.

2.1.3. The Decomposition Argument.

Fodor does not, of course, regard himself as committing a category mistake in explaining intentional activity in terms of the prior operation of the thinking mind. In particular, he can agree with Ryle about the misleading aspects of Cartesianism. Firstly, it is not essential to Fodor’s story that the relevant intellectual processes are accessible to consciousness. Secondly, Fodor also hopes to avoid the regress argument by positing a progressively less complex hierarchy of intellectual processes which bottom out in "elementary operations." An elementary operation might be the activation of one node in a network of neurons or the flow of current in a circuit. These elementary operations can be combined in a progressively more complex hierarchy of operations to produce representations and to link these representations with behaviour. This is the Decomposition Argument. In Fodor’s story, the rules for combination and recombination culminate, at the top level, in the sort of rules invoked by intentional psychology.

At the bottom level of the hierarchy, propositional knowledge is not involved. The rules the system follows are rules like "if - then " or negation. The bottom level of the system does not have to judge whether a proposition is true, it merely responds automatically to stimuli. At the bottom level the system is like the pixels in Dennett's game of Life which switch on and off
according to whether or not adjacent pixels are illuminated. On Fodor's story intentional understanding is realised by systems which, at the bottom level, "understand" nothing.

On one way of understanding Fodor's claim this must be true: intentional understanding depends on the human brain which is a system of millions of cells whose local connections are with their immediate neighbours and governed by simpler and different rules than the principles of intentional psychology. Yet this brain represents the propositions believed by the agent of which it is a part. On this reading of Fodor, or any other neo-Quinean who wants to claim that human brains represent the theory of intentional psychology and use it to govern their behaviour, there is no real conflict with the Rylean position. What Ryle describes as practical knowledge of propositions will turn out to be realised by information flow at the microscopic level.

Fodor, however, has a stronger claim in mind: that the theory of intentional psychology is represented as a system of symbols together with rules for their combination and recombination. This is of course the position that Ryle denied was possible. However, Ryle denied it only for the explicit conscious or unconscious tokening of rules and symbols, not for symbols reconceived of as abstract descriptions of processes which can be progressively decomposed in the way Fodor describes. Whether Ryle's arguments work against accounts of mental representation which employ strong versions of the decomposition argument depends very much on how those versions explain what it is to mentally represent a rule or a symbol.

2.1.4. The Neutrality Thesis

While I take no position on this question it is important to set out some of the issues because I wish to maintain later in chapters three, four and five, that the child who acquires the concepts of intentional psychology can do so without
thereby being committed to any account of the nature of the mental representations involved in the tokening of propositional attitudes. So I need to show why someone who ascribes propositional attitudes is thereby not automatically ascribing internal tokening of symbols which correlate with those propositions and rules which correspond to the inferential relations among propositions.

Consider again Dennett's example of the game of life. Imagine that the screen produces this series of patterns, a fish, followed by a duck-rabbit, followed by an elephant. I am happy to allow that the computer represents this proposition: "follow the fish with the duck-rabbit." However the computer does not do so so by representing that rule or by representing symbols for a fish or a duck-rabbit. Any rules represented are at a far more basic level, describing interactions between individual pixels.

An alternative way to produce this result might be to program a computer with a symbolic code which has symbols for a fish, a rabbit, and an elephant and a rule for their succession. Of course this code will be implemented in such a way that the relevant pixels are illuminated, according to the program. In this case a system of rules and symbols are realised in the manner suggested by the decomposition argument.

The point is that where we attribute the capacity to represent a proposition in order to explain correlations between input and output, it does not automatically follow that the system works by manipulating a symbol or symbols corresponding to that proposition. In the example the systems represent the same proposition: "follow the fish with the duck rabbit." In the first case, without using rules and symbols, in the second case, using rules and symbols.

The point of this chapter is that when we explain other people's actions in terms of the propositions they represent we are not thereby endorsing any account of the actual processes involved. Our brains might operate like the game of Life, like classical computers, or in some other manner yet to be
discovered. Since intentional psychology does not purport to state any hypotheses on this subject, the practice of interpreting other's actions does not carry with it any implicit endorsement of any theory about how propositions are represented. This is the Neutrality Thesis.

The neutrality thesis is important, because if we remain neutral on the nature of representation then we will not rule out any particular account of the nature of the representations involved in understanding other people. In the next two chapters I argue that developmental psychologists have assumed an inadequate model of representation with the result that their account of the abilities involved in intentional interpretation is too narrow. With this in mind, consider the Rylean arguments against the Cartesian version of the Rules and Symbols account of representation. The interest of these arguments is, firstly, the support they give to the neutrality thesis by arguing against one particular account of mental representation, and, secondly, the way they bring into focus the variety of abilities and capacities involved in intentional understanding.

2.2. Salience and Intentional Activity

Recognising the force of Ryle's argument that intentional activity is the exercise of a skill, acquired through learning, which is innovative and adapts to changing circumstances, early practitioners of artificial intelligence set out to build systems with these types of ability. They did so using a method of which Ryle would have disapproved: programming computers. In doing so they ignored premise 3 of the regress argument: that any activity produced by the mechanical following of instructions is not intentional. Their intention, however, was to refute premise 3 empirically. Their acid test, with which Ryle would have agreed, is the acquired ability to converse fluently in natural language. A system which could learn to do this should qualify, on Ryle's test, as an intentional system. Of course Ryle's argument was an a priori one, but it
is safe to assume that the early practitioners of artificial intelligence had either read Quine or had no patience with such conceptual sleight of hand.

The philosophical responses to the early attempts of artificial intelligence to design such a system supplement Ryle’s account of intentional activity. Where Ryle pointed out that any propositional knowledge mobilised by an agent must be salient and then went on to claim that this requirement could not be met by mechanical application of rules, Dreyfus⁹ tried to develop an account of the nature of salience itself in order to explain why purely formal systems could not achieve it.

According to Ryle, the ability to speak a language is a matter of having the right sort of flexible dispositions to respond in conversation and to generate new sentences as appropriate. Ryle, quite rightly, says that we acquire these dispositions in the course of our early, everyday linguistic interactions. Ryle, however, gives us no clue as to how this is accomplished, except to say that it is a skill and, like all skills, acquired through practice and training and practised intelligently and flexibly. He also tells us that the regress argument shows that any kind of intentionally practised skill cannot be explained as the enaction of prior knowledge that or a theory of the activity. How then do we acquire the ability to, as Wittgenstein puts it, "go on in the right way" which is definitive of intentional activity?

There is one solution to the problem, which Ryle would not countenance, and that is that from one’s limited exposure to language one acquires, or activates an innate set of, rules for understanding and producing meaningful sentences. The obvious move for such a theorist to make is to say that the rules are tacitly represented in the cognitive systems that subserve speech production. In such a situation the speaker would be unaware of the rules which governed her spoken output although she would, nevertheless, be guided by them in her linguistic interactions.

⁹ And others such as Haugeland. Like Dreyfus, Haugeland is drawing on the work of Heidegger on preconceptual understanding of the social environment. See especially (Haugeland 1979)
Chapter 2

When Ryle mounted his anti Cartesian campaign, theories of internally represented generative grammar, or computational theories of artificial intelligence, were not part of the Cartesian repertoire. His arguments were directed against the explanatorily vacuous procedure of explaining an intentional activity by treating it as the product of the same activity internalised. Computational theories, however, do not merely internalise a language or a theory, they purport to explain it in terms of a different, more fundamental theory or set of rules for performance, underlying the public activity.

This computational theory underscores Good Old Fashioned Artificial Intelligence. GOFAI theorists operated within a classical computing framework, one in which symbols are manipulated according to syntactical rules which are specified in the program. Thus they programmed their computers with a translation of the grammatical rules of English and the contents of a dictionary. In essence they were attempting to provide the computer with the ability to generate, according to the program, all the propositions which an English speaker knows about English. They trusted that, as a result, the computer could hold its own, initially in restricted situations, but eventually in full blooded conversation, with English speakers. They were hoping that the computer would exhibit the same dispositional profile as any other English speaker.

Ryle would have objected to GOFAI for the negative reason that it attempts to reproduce intentional activity by complex, mechanical, rule following. The Berkeley Heideggerians, Dreyfus and Haugeland, criticise GOFAI for essentially the same reason as Ryle, that it attempts to locate meaning and intentionality in the prior application of theory rather than the multitrack dispositions which constitute that activity. However, their attack on GOFAI adds to Ryle's account a story of how it is that we acquire the multitracked and open-ended dispositions.
Human speech is governed by pragmatic and contextual constraints on what information can be conveyed by an expression. The classic cases are pronouns and demonstratives. The sentence "Mary bought Jill a car but she smashed it" is an example. Who smashed the car, Mary or Jill? This is a basic example which depends on establishing which proper name binds the pronoun. Normally, in speech situations, the conversants have sufficient background information available to disambiguate the sentence. But this additional information is supplied by the context in which the sentence is produced, not by the sentence itself.

These contextual constraints on what can be conveyed in a given speech situation fall under the general heading of relevance. Considerations of relevance multiply exponentially as soon as one considers continuous communication in complex sentences. Such communication requires a constant sensitivity on the part of the participants to relevant information which is not expressed by the sentences themselves. A relatively simple expression like "She is a Napoleon" introduced during a discussion of a colleague's relation to her supervisor introduces an enormous set of relevant considerations. Does "she" refer to the professor or the colleague? Does the speaker mean that the supervisor, or the colleague, has a dictatorial manner, or that she is silently plotting revenge, like Napoleon on St Helena? And so on.

Relevance theorists argue, plausibly, that the importation of this type of background information is essential to communication. The GOFAI theorist quickly has to acknowledge the force of the relevance considerations when she designs a system to communicate with other language users. Initially she will try and encode the relevant background information, together with its rules for introduction as the conversation continues. The task of the GOFAI theorist is to write down the syntax and semantics of a language plus the additional relevant background propositions which must be known by a participant, together with a
set of rules which would allow the system to determine which background propositions are relevant. 10

The interesting thing about Dreyfus' objection to GOFAI on the ground that it cannot deal with Relevance, is that he purports to explain the difference between acting mechanically and intentionally. The difference is that rules and theories abstract away from the lived experience of the agent. What is communicated by language (information, thought, the content of experience) cannot be captured by a system of rules and symbols abstracted from the situations in which language is employed. As Dreyfus says:

This still leaves the important question of how human beings manage to engage in purposive behaviour. The traditional view, accepted by GOFAI, has been that they use their theory of the domain in question to work out a plan for accomplishing whatever it is they are trying to do. But rather than suggesting that people store vast numbers of facts and then plan how to use them, Global sensibilities (or the imagination thereof) determine situational relevance because our world is organised by these preconceptual meanings. It is in terms of them that everyday objects and events are experienced as something. Our everyday coping skills and the global familiarity they produce determine what counts as facts and the relevance of all facts and so are already presupposed in the organisation of frames and slots GOFAI uses for representing these facts. That is why human beings cope more easily and expertly as they learn to discriminate more aspects of a situation, whereas, for data bases of frames and rules, retrieving what is relevant becomes more and more difficult the more they are told.11

This is a manifesto, not an argument, and Dreyfus goes on to support his claims, based on these three considerations,

1. the primacy of lived experience over a theoretical account of that experience, at least until science can account for the way in which that experience presents itself to us.

10 Fauconnier has a discussion of relevance in terms of government and binding of anaphora in (Fauconnier 1985) which also includes an original theory of the cognitive processes underlying the understanding of language in context. 11 Dreyfus, 1992, pp xviii-xxix
2. the role of perceptual judgement in that experience.
3. the link between perceptual judgement and the actions which constitute practical ability.

2.2.1. The Link Between Perception and Action

1. does not look as if it can function as part of the argument because the issue is precisely whether lived experience is, or is not, lived via the application of a (tacit or innate) theory. 2. and 3. are the essential ones. The key idea is that the act of perception itself narrows down the range of possible responses and consequent actions for a perceiver. If I see a spade, for example, I see a digging implement, or part of a conceptual art installation, depending on the milieu in which I have been socialised. I do not see the other myriad possibilities for which the object could conceivably be used. For instance, I do not see a portable mirror, although a spade could easily be polished up and stood against the wall upside down to perform exactly that function. The idea is that we perceive the world in the same way as those with whom we share an experiential background of education, culture, language or geography. Perhaps a maritime culture, which traded and fished but never farmed, would see a spade as a paddle. Our concept of 'spade' could be explained to them but it would not be effortlessly assimilated in their practice and culture in a way which allows its artefactual role to be taken for granted amongst a community of gravediggers or gardeners.

Familiar objects like the spade/paddle are perceived automatically as artefacts for a certain type of use by the culture which relies on them. According to the Berkeley Heideggerians, the explanation of relevance lies at the level of a pretheoretical sensitivity to patterns of activity. In perceiving her surroundings the acculturated language user automatically brings to bear the appropriate background. That is why the expression lebenswelt is used to
describe the background. A life world is the world as experienced by those who act in it. It is perceived in terms of the scope it offers for action and interaction with others against the background of familiar environmental patterns and structures. It is not first perceived as "raw" data of experience which are then tacitly processed according to the rules for applying concepts which govern that lifeworld.

An important part of relevance is the excluding of possibilities. For example, excluding the possibility that a spade is a paddle, a weapon, a mirror or an ornament. The Heideggerians emphasise that, in familiar circumstances, the possible ways of incorporating familiar objects into our activities are already limited by our previous experience of the structure of our environment and the patterns of human activity which it sustains. So our way of life already forecloses certain possibilities. Since that way of life is second nature to us, we perceive the world as the limit of the range of activity possible for us. Persons who are acculturated into different patterns of activity will perceive their environment as a different range of possible constraints on activity.

So too with speech and writing. Persons who have acquired their language, or participate in dialogue, against the same lebenswelt understand language with the same set of preconceptions in mind. They have no need to make explicit the relevance considerations which hinder communication with outsiders. The point which Dreyfus is emphasising is not that those preconceptions cannot be articulated as a theory, they can. However they are acquired, internalised as second nature, through mastery of a practice of habitual action and response within the social milieu.

We can summarise Dreyfus' point by saying that practical ability is best explained in terms of the immediate link between perception and activity. Perception, in the sense Dreyfus is using it, is not just the registering of sensory stimulation, but the basis for a judgement that the world is in a certain state. The agent makes those judgements according to the way the world constrains her possible activities, activities which are themselves constrained by a form of
life to which she is acculturated. That process of acculturation is a matter of recognising the patterns of activity in the world around her. According to Dreyfus, none of these cognitive tasks, perceptual recognition or consequent decisions to act in one way rather than another involve mental representation of an articulated conceptual scheme. Practical ability is grounded by pretheoretical perceptual experience.

I suspect that many GOFAI theorists, or cognitive scientists of any persuasion, will grant that Dreyfus correctly describes the way our ability to negotiate familiar milieux feels to us. It seems to us that we perceive the world directly, not via the mediation of a theory and that we act on the basis of those perceptions without first submitting the information they give us to any sort of decision procedure. This is especially so when we are having a conversation or riding a bike, procedures which are quintessential cases of practical ability. However what those theorists will not concede is that the way it seems to us is the way it is with us.

2.2.2. Lebenswelt or Conceptual Scheme?

Philosophers have long been interested in the relation between perception and judgement and in the way perceptual judgement depends on the organisation of "raw" perceptual material. Traditionally they have described this process of organisation as the application of concepts, where a concept is something like a rule for processing the flux of experience. Kant famously tried to deduce, a priori, from the nature of our experience, the basic concepts which must structure that experience: "it is possible to show that pure a priori principles are indispensable for the possibility of experience and so to prove their existence a priori." (B5)

Since Kant, others have tried to show that our perceptual experience can be conditioned, not just a priori or by the structure of our perceptual apparatus but sociologically. They claim that the difference between seeing the same
object as a spade, a mirror or a paddle is a matter of the conceptual scheme within which we learn to recognise objects. The conceptual scheme is a theoretical articulation of the *lebenswelt* against which we make perceptual judgements. Such theorists do not dispute the Dreyfusian claim that perceptual abilities are not experienced as the application of a theory. Rather they claim that that practical ability, which seems immediate as we exercise it in linking our actions to our perceptions, actually depends on the internalisation of a set of concepts, embodied in social practice, which make up a theory.

The essential idea, lucidly expressed by Paul Churchland, is that the terms which we use to report and record the impact of perceptual information on our sensory systems get their meanings from their role in a network of theory. As a result we acquire an "observation vocabulary" in which to record and communicate the results of our perceptual sensations. The observation vocabulary constrains the "objective intentionality" of perception — its reliability as an indicator of the presence or absence of features of the environment. The observation vocabulary ensures that perception is focused on those aspects of the environment which are salient within the observation vocabulary. It follows that were that vocabulary to change radically we would become sensitive to different features of the environment without the need to modify our perceptual equipment.

Churchland's example is of a society whose observation vocabulary changed to that of physical theory. In such a society, purely in virtue of the language used to report and convey the content of experience, people's perceptual judgements would track features of the world to which we are presently not attuned.

It is important to try and appreciate, if only dimly, the extent of the transformation here envisaged. These people do not sit on the beach and listen to the steady roar of the pounding surf. They sit on the beach and listen to the aperiodic compression waves produced as the
coherent energy of the ocean waves is audibly redistributed in the chaotic turbulence of the shallows... They do not observe the western sky redden as the sun sets. They observe the wavelength distribution of incoming solar radiation shift toward the longer wavelengths (about $0.7 \times 10^{-6}$ m) as the shorter ones are increasingly scattered away from the lengthening path they must take as terrestrial rotation turns us slowly away from their source. (A transparent consequence of this situation is that the wavelength distribution will swing visibly away over to the other extreme after our line of sight to the sun is cut off and the only visible radiation to reach us is at those shorter wavelengths (about $0.45 \times 10^{-6}$ m) scattered around the curve of the earth's surface. That is to say the western sky will be 'bluest' some time after sundown.\textsuperscript{12}

For Churchland, human perception is a process like the recording of information by a scientific instrument. Both systems are designed to be sensitive to certain features of their environment, to communicate the presences of those features and fluctuations in their levels. To do so they need an interpretation function which converts the causal impact of the environment into a format suitable for expression in the observation vocabulary in which output of the system is expressed. The choice of interpretation function depends on which features of the world are salient for us.

Interpretation functions are not sent from heaven, nor are they written \textit{a priori} in the output state of our measuring instruments. They derive from, or reflect, our current understanding of the world.\textsuperscript{13}

That current understanding, articulated and systematised into a theory, is the conceptual scheme.

Dreyfus does not claim that there are no conceptual schemes, because he would certainly agree that the set of presuppositions implicit in a cultural practice, including scientific investigation, can be made explicit and articulated.

\textsuperscript{12}(Churchland 1979 p29)
\textsuperscript{13}(Churchland 1979 p39)
into a theory. This type of articulation is, after all, the project of the social sciences. However he denies that the interpretation function operates by representing the concepts which make up the conceptual scheme as a system of rules and symbols.

Interestingly, Churchland also rejects the formal, rule governed, conception of mental representation which underpins GOFAI. His work on conceptual schemes is designed to show, not that we think via manipulation of the concepts they contain, but that our interpretation functions are adapted to our environment and observation vocabulary. Exactly how interpretation functions operate to adapt our perceptual responses to the environment is presently mysterious. His explanation of the way they do not function is not so different from the argument Dreyfus gave against GOFAI and Ryle gave against Cartesian introspection.

Now it plainly will not do to suggest that each of us "sits behind" his personal battery of measuring instruments (sense organs), observes their sensational outputs, and uses an interpretation function in formulating his perceptual judgements. For one thing, this misrepresents badly the psychological facts of normal perception. For another, it clashes with the fact that children acquire the ability to observe and describe the world in great detail before they have any significant or explicit awareness of the richness of their sensational life, or even of its existence...

But if it is plain that perception does not involve the use of an interpretation function in the explicit manner suggested, it is equally plain, insofar as our conceptual responses to our sensations do display determinate and identifiable patterns, that we embody or model a set of interpretation functions, functions implanted in childhood as we learn to think and talk about the world in the language of our elders. 14

The issue is how we embody or model that set of interpretation functions. Ryle and Dreyfus have argued that we do so through our lived

---

14 (Churchland, 1979 p39)
experience, manifest as a set of perception-based practical abilities. Whether or not these arguments tell against the idea that practical, or ability based knowledge can be reduced to the manipulation of rules and symbols depends on how one understands "rules and symbols." The arguments are successful against the Cartesian versions described at the beginning of the chapter. However on the weak reading of what it is to mentally represent rules and symbols, all can agree that we mentally represent the propositions and principles of intentional psychology. After all, Churchland and Dreyfus, who both argue against the rules and symbols approach agree that we embody or model a set propositions.

What Churchland describes as the acquisition of a conceptual scheme and Dreyfus the ability to negotiate the *lebenswelt* are not necessarily conflicting accounts of the acquisition of intentional abilities. Churchland agrees with Dreyfus that purely formal systems which seek to encode conceptual schemes have, thus far, been unsuccessful in replicating intentional activity such as language use.

In a nutshell, the problem is that to understand natural language at the human level requires that one have an overall knowledge of the world that is comparable to what a human possesses and we have not solved the problem of how to represent and store such a knowledge base in a fashion that would make access and manipulation feasible. Related to this is a deeper problem. We have not even solved the problem of how such global amounts of knowledge can even be acquired.15

Churchland's account of the acquisition of the knowledge base required for intentional activity and the way it is applied, is also similar to that of Ryle and Dreyfus.

---

15 (Churchland 1988 p119)
What begins as a set of stimulus - response patterns gets progressively articulated and refined as the children acquire from their elders an appreciation of the rich fabric of entailment relations characteristic of the theoretical framework as a whole. As this appreciation grows, so does their ability to capitalize on those otherwise pointless patterns of linguistic response implanted initially. Their mature observation judgements are bursting with systematic implications.\(^{16}\)

Churchland explains intentional responses as the extension, through learning and training, of capacities which are initially merely rote or mechanical. Where the child initially perceived a situation in terms of a single or narrow range of possibilities for action or response, after exposure to the social milieu they perceive a wider range of appropriate responses. Dispositions which started out as single track become, through training, multitrack and open ended. Churchland describes this process as a widening of the range of possible inferences which the agent can make as the result of her perceptions of a situation. Inferences grounded in the rich fabric of entailment relations characteristic of the conceptual scheme as a whole.

Entailments and implications are relations which hold between the propositions which make up the conceptual scheme. Insofar as behaviour fits the conceptual scheme, the agent can be said to model or embody the structure of the scheme. Churchland is not suggesting that she explicitly represents, as an articulated set of propositions, any of the possible uses of a spade, or propositions about agrarian culture in general, in order to acquire this sort of familiarity. But he would maintain that a certain set of propositions which make up the agrarian conceptual scheme are true of her, and their inferential relations can be used to explain her behaviour. So there can be a true theory of agrarian society and social behaviour within that society without it being the case that anyone in that society explicitly represents that theory. Perhaps this is obvious. Theories which systematise and model the behaviour of individuals

---

\(^{16}\)(Churchland 1979p29)
do not depend for their truth or falsity on whether or not the individuals themselves know the theory and use it to govern their behaviour. Consider the case of language. A linguist provides a grammar, $G$, for a language, $L$, which specifies rules for the production of meaningful sentences within that language. Provided that all sentences, recognised as correct by native speakers, can be described by the grammar the linguist can rest. Her theory is correct. It is not a requirement on the correctness of $G$ that all native speakers explicitly represent $G$ and produce sentences according to it. $G$ is correct if it is true of the sounds and marks made by language users, not in virtue of whether the set of propositions which comprise $G$ produced by the explicit application of $G$. This echoes Dennett's point about the patterns in the game of Life. One might become a very good predictor of the succession of patterns based on observation of regularities without purporting to describe the nature of the process by which the patterns are produced.

Ryle, Dreyfus, Dennett and Churchland point out that there is no a priori necessary connection between the structure of a theory of intentional activity and the structure of the internal processing (present to consciousness or not) which causes that activity. Consider $G$ again. No doubt there is a description of the mental processing, the neural activity, by which speakers parse and produce sentences. This description can be systematised and generalised into a theory, $G^*$ of the "deep structure" of $L$-understanding. In this sense $G$ is peripheral to $G^*$. As Ryle was at pains to point out, someone who has learnt $G$ can exploit it to produce grammatical sentences, perhaps applying it in foro interno before speaking aloud. This ability to consciously exploit the peripheral

---

17This is the weakest way to put the claim. If psychological states are natural kinds, and something like the Kripke-Putnam theory of reference for natural kind terms is true, then an easy way to put Churchland's version of the error theory for folk psychology is to treat him as denying a necessary a posteriori connection. Many of Churchland's remarks suggest that this is the correct way to understand him, and certain of his targets, such as Fodor, have, at one time or another, been happy to define their enterprise as a search for the natural kinds which correspond to the terms of Folk Psychology. The commonsense functionalists, like Jackson and Lewis, have a different theory of reference for theoretical terms so Churchland's argument cannot touch them unless he can first convince them of the truth of his theory of reference.
theory, however, depends on the processing explained by \( G^* \), it is not evidence for the structural similarity or identity of \( G \) and \( G^* \).

Dennett and Ryle explain the theoretical appearance of intentional explanation in the same way. It is a peripheral theory of intentional activity which does not depend for its truth on whether or not it limns the "deep structure" of mental processing. That deep structure will be explained by the theory of neurophysiology or cognitive science. There is no reason to suspect that the relation between the \( I^* \) theories of these disciplines and the \( I \) theory of intentional explanation would be constrained by structural similarity, any more than we should expect that structural constraint to apply to the case of \( G \) and \( G^* \).

So, to use Churchland's phrase, an agent can *embody* or *model* a peripheral theory of intentional activity, even though the agent does not explicitly or tacitly understand any of the propositions which make up the peripheral theory. The agent does, however, have practical knowledge, knowhow, exhibited in a set of multitrack dispositions whose relations are displayed by the inferential relations among the sentences of the peripheral theory.

These themes will be echoed in later chapters. In particular in chapters three and four I wish to show that the fact that children learn to understand and apply the concepts of intentional psychology does not automatically entail that they internally represent those concepts as a set of rules and symbols in the strong sense. Of course this is not to deny that children internally represent the propositions which make up intentional psychology but to point out that the neutrality thesis reminds us that the relation between any \( I \) and \( I^* \) theories is a matter to be determined by empirical investigation.

In chapter six I make a similar point about the acquisition of cultural concepts, the domain in which Dreyfus mounts his Heideggerian argument. I argue that we can explain the acquisition of culture in the same way as we explain the acquisition of the concepts of intentional psychology. Perhaps as
some theorists claim, concept formation in the two domains exploits the same neural architecture. The claim is not *prima facie* implausible: most of the interpretative effort involved in understanding cultural communication is a matter of determining and responding to the beliefs and desires of one's fellows. So the neural architecture the young child uses to represent the concepts of intentional psychology may well be employed in the understanding of cultural communication. The neutrality thesis, supported by the game of Life example, reminds us, however, that the fact that a child represents certain concepts does not entail the truth of any hypothesis about the nature of that architecture.

### 2.3. Peripheral Theories and Practical Knowledge

Ryle and the Berkeley Heideggerian drew a distinction between the sort of knowledge which grounds skills, or coping abilities exercised within a specific cultural or cognitive context, and intellectual knowledge of propositions or theories composed of propositions. While this is a useful preliminary distinction I am less interested in the distinction between *knowhow* and *knowing that*, than in a distinction between two ways of understanding propositions, a distinction which Philip Pettit has called the distinction between practical and theoretical understanding. In particular, I intend to make use of the concept of practical knowledge of a proposition or set of propositions in my account of the acquisition by young children, of intentional theory. My claim will thus be not that young children gain *knowledge how* to apply intentional concepts rather than *knowledge that*. Rather I will claim that young children gain practical knowledge of the truths of intentional psychology rather than intellectual knowledge of those truths. While intellectual knowledge and explicitly represented *knowledge that* are similar concepts, practical knowledge is not quite the same thing as *knowledge how*, because practical knowledge is knowledge of a set of propositional truths, whereas, as we saw previously, *knowledge how* is not
truth evaluable. It turns out that practical knowledge is something one has in virtue of a set of knowhows.

The chess player, musician, reasoner or speaker of a language who learn by example and practice rather than formal instruction all exhibit this type of practical knowledge of the theory of their disciplines. It is practical knowledge of the theory in the sense that the theory articulates, in terms of a system of interrelated concepts, knowledge which they possess through their abilities and capacities, and which guides the exercise of those abilities and capacities.

Philip Pettit, in his discussion of the distinction between practical and theoretical understanding, talks about believing propositions or possessing concepts in the intellectual and in the practical modes. As he says, to believe a proposition in the intellectual mode is to be able to represent the elements of the proposition as part of the process of judging whether it is true. This requires the ability to abstract, to make the proposition an object of thought. We could apply Pettit’s distinction to the case of musical knowledge in the following way. A musician who understands, in the intellectual mode propositions which describe her understanding of music, would be able to say that what makes a minor key is the flattening by a semitone of the third note in the scale commencing at the tonic note. No doubt such a person would be able to compose music and articulate their response to music using these concepts. Similarly, when the leader of her ensemble asks her to try a piece in the minor key rather than the major she is able to make the adjustment by representing the concept and adjusting her playing accordingly.

However, it is also possible that a person understands music solely in the practical mode. Such a person might have learnt music by listening to, and playing along with, other musicians but without ever having the concepts explained to her verbally or being taught how to read music. She has no intellectually represented concept of a tonic scale, a minor or a major key. Nonetheless, when the other musicians in her ensemble start playing in a minor

18 (Pettit 1994)
key, she too flattens the third note in the tonic scale, when they switch back to a major key she reverts back to playing normally. She has all the dispositions of the other musicians who have an intellectual understanding of music but she is unable to make the general statement "All minor keys flatten the third" because she doesn't possess any of these concepts intellectually. Nonetheless she is able to play the music in concert with the other musicians who have learnt in the more conventional manner and who have an intellectual understanding of the relevant concepts.

Intentional activities require knowledge: of music, rules of language or logic. Intentional explanations therefore attribute agents with propositional attitudes, of belief, knowledge etc to the propositions which encode the relevant musical, grammatical or logical theory. Because theory deals in universal propositions ("all minor keys flatten the third", "all people who fear $p$ desire, ceteris paribus, not $p$") it is natural to explain behaviour which is flexible and which extends from limited experience in terms of the application of a theory made up of a set of such propositions. This is what Churchland means by the "rich network of entailment relations" embodied in the activity of an agent to whom intentional theory applies.

He does not mean, however, that the agent represents the theory and consults it when in doubt as to what to do, or follows it like Fodor's homunculus. Rather he means that the dispositions whose structure the theory displays are produced as a result of practical knowledge of the propositions which make up the theory.

2.4. Practical Knowledge and Rule Following

The distinction between practical and intellectual belief gives a new perspective on a notorious problem for dispositional accounts of mental attitudes, a problem which I discuss in chapter six and which follows from the brief discussion of rationality in chapter one. Recall that Mackie suggested that a
purely dispositional account of mental states would have trouble explaining the rationality of behaviour, the ability to think about and revise one's beliefs and to deliberate about courses of actions. On the dispositional account how could we distinguish between behaviour which is rational, and invites intentional explanation, and that which is not, without being forced to postulate some interior process of deliberation or reasoning? The practical/intellectual distinction suggests a possible reply: activities which constitute practical knowledge of an intentional theory are intentional.

If we allow that a theory constitutes a rule for an activity then the concept of practical knowledge allows us to say that an agent can know and follow the rule for an activity even though she does not intellectually represent the rule and consult it. This approach, if it is successful, will be a partial response to Wittgenstein's challenge to state the difference between blind, brute, behaviour which happens to coincide with a rule and intelligent purposeful activity which can be said to "follow" the rule. I do not present it as an answer to the rule-following problem, rather I use the discussion of the rule-following problem to refine the account of the distinction between practical and theoretical knowledge, and argue that it is plausible to attribute knowledge of some truths on the basis of dispositions. While this is not an answer to the rule following problem, it has similar features to solutions to that problem proposed by Pettit, Goldfarb and McDowell, all of whom invoke elements of the dispositional account in their replies to Kripke's sceptical objections. 19

19 The interpretation of Wittgenstein's remarks on rule following is a problem whose complexity is well beyond the scope of a single thesis. In what follows I merely try and show that it is plausible, for some instances of practical knowledge, of music and of ethical principles, to allow that they are instances of practical knowledge of rules for the conduct of the relevant activity: cognitive activities which do not depend on explicit representation of the propositions true of the activity. Such a view is consistent with the view that Wittgenstein was not mounting a sceptical attack (see his remark at 201 of the *Philosophical Investigations* "there is a way of grasping a rule which is *not an interpretation* but which is exhibited in what we call "obeying the rule" and "going against it" in actual cases") but directing our attention to something like the way we learn our patterns of behaviour against a shared background of experience. McDowell is a consistent articulator of something like this way of understanding Wittgenstein, and, perhaps a defender of the dispositional thesis in something like the Aristotelian form. See (McDowell 1982) and(McDowell 1984) especially pp330ff. Others who have defended similar versions of a non sceptical account of rule following are (Goldfarb 1985) and (Pettit 1990)
Wittgenstein is sometimes represented as a sceptic about possibility of rule following for any, or all, of the following reasons:

1. There are no rules, if by rule we mean a universal proposition which can be consulted to guide behaviour. The Relevance considerations (also relied on by Heidegger and Dreyfus) show that rules for open ended flexible behaviour cannot be codified in this way.

2. The regress argument shows that even if there were such rules, consulted in the production of behaviour, such consultation is neither necessary nor sufficient for intentional behaviour.

3. Following a rule cannot be a matter, in the absence of explicit representation of the rule, of following one’s inclinations, where inclination is something like a qualitative sensation which one associates with the object of activity and prompts one to act appropriately (the private language argument, which I shall not mention again).

4. Following a rule cannot be solely a matter of having a certain set of dispositions for two reasons:

   (a) dispositions can be mistaken. (Although, strictly speaking, dispositions are not the sort of thing which can be right or wrong. Nevertheless, one’s dispositions might lead one away from a rule, as in the case of a musician who goes subtly out of time without realising it.)

   (b) a finite set of dispositions does not determine a single rule. A finite set of dispositions can be interpreted to coincide with a multiplicity of rules.\(^{20}\)

All of these considerations are good reasons for not reductively identifying rule following with any of the candidates Wittgenstein rules out. Those who treat Wittgenstein as a sceptic tend to regard him as providing an alternative conventional solution. The idea is that since we cannot consult our inner sensations or representations or rely on our dispositions to tell us whether or not we are following rules, or which rules they might be, we allow

\(^{20}\) (Kripke 1982) develops the sceptical attack on the possibility of equating rule following with a set of dispositions, see eg p41.
the community to determine, as a matter of convention, whether or not activities are done correctly. 21

However Wittgenstein is not a sceptic about rules and rule-following behaviour. He does think that we follow rules, that the patterns of our dispositions are the result of our cognitive efforts. However those cognitive efforts should not be seen as the construction and consultation of intellectual knowledge, of a theory which encodes the relevant rule(s) but of the development of practical knowledge in the same way as the musician develops her musical ability.

Consider a possible response which a musician, part of a string quartet, might offer to a Kripkensteinian sceptical analysis of her practical ability to play by ear the second violin part of a Beethoven string quartet:

Firstly, the idea that the activity is not rule-governed is a non-starter. At every rehearsal the players reproduce the same piece of music. They stop when they hear a mistake and go over the section until the group sounds right. They are rehearsing in order to habituate themselves to the production of the pattern of sounds, but it is not merely rote behaviour, because it is produced on every occasion with the intention to get it right and to correct mistakes.

Secondly, the musician who learns her part by ear doesn't need to read the score in foro interno or pay attention to some peculiar internal sensational prompting (perhaps "hearing" the music in her imagination). Even if she did so, as Ryle observes, such internal episodes would not explain her ability to play the music: they would merely accompany that ability. So the musician can agree with Wittgenstein and Kripke that following a rule cannot be a matter of consulting an internal sensation or representation, without being forced to adopt the sceptical conclusion. She can say that she follows the rule in virtue of her practical knowledge of it. Of course, in so doing, she needs to respond to the sceptical objections to a dispositional account of rule following.

21 (Kripke 1982pp77-78) "All that is needed to legitimize assertions that someone means something is that there be roughly specifiable circumstances under which they are legitimately assertible, and that the game of asserting them has a role in our lives"
It is true that, considered in isolation, the dispositions of the musicians might not determine any particular rule. When practising by themselves they might misremember their part or develop an unnoticed technical flaw which puts them subtly out of tune or out of time. For any single instance of playing, however, the terms "out of tune" or "out of time" are not really applicable. The question arises "out of what tune ?" A single set of dispositions cannot thus determine a rule. If, however, the single set of dispositions is produced with the intention to conform with a larger pattern this problem dissolves: consider again the musician practising in isolation who makes a mistake. Perhaps, as Kripke suggests, the concept of mistake is not applicable here. However, when she rejoins the ensemble her mistake will be obvious through the dissonance it produces and the reactions of her colleagues. She will then fix the mistake. If the ensemble had no dispositions to amend their performance in this way their private practice would not count as following the rule for playing that particular piece. This shows only that rule following behaviour requires the dispositions to recognise and correct mistakes by submitting oneself to the judgements of others engaged in the same activity, as well as the basic dispositions counted as correct according to the rule.

Finally, imagine the second violin rehearsing by herself. As she does so she acquires and strengthens a set of dispositions. She believes, in the practical way, that she is producing these dispositions in the attempt to play the second violin part of a specific string quartet, (whose score constitutes the relevant rule). Of course it is possible that she is following the rule for some other performance, perhaps a rule which says that after the five hundredth performance she plays the last movement backwards from then on.

Considered in isolation, her rehearsals alone cannot determine which of these possibilities is she is realising in her performance. However the way to determine the matter is to have her play with the rest of the musicians. If they

---

22 I see no objection to allowing " others " to stand for former and later stages of the same self. For the purposes of rule following one can submit oneself to the adjudication of one's memory or subsequent inclinations. Of course, one is more likely to be mistaken in this case, but that only shows that the private following of rules is a fallible affair.
baulk at her five hundred and first, and subsequent performances, ask her to stop and listen to them playing the last movement backwards, then the alternative rule is the one she has been working towards. She just has not played enough to realise it. This doesn't mean that it is indeterminate which rule she is attempting to follow, just that, over the long haul, proficiency at rule following depends on a continued willingness to stay involved with other participants in the practice.

All these replies invoke the community of rule followers and make a background of shared participation essential. However the community is not invoked to provide a stipulative or conventional standard in the face of otherwise insuperable sceptical problems. Rather, the community provides the essential background for the acquisition of the practical knowledge of the relevant rule. The dispositions which make up the agent's practical knowledge are learnt and expanded within a wider community whose practices carry information about the standards which govern their activities. Anyone following a rule it is a musician rehearsing for a performance. Rehearsing a piece of music is a matter of acquiring practical knowledge: ie becoming aware of standards for the performance of an activity, trying to meet them, recognising shortfalls or errors and revising and correcting one's attempts. All of these conditions can be met without the intellectual representation of the rules for performance of the activity, provided that, as in the musical case, the skill is acquired against a background of joint participation with others competent in the skill.

2.4.1. Aristotle on Ethical Knowledge

Another example of the relation of practical to intellectual knowledge is contained in Aristotle's account of the acquisition and practice of ethical

\[23\text{As would be the case if the participants agreed on the following: "there are no rules, or even if there are, we can never know them, so let's agree to follow the majority"}\]
understanding. Aristotle gave an answer to a question Socrates considered in the *Meno* (70a): "can virtue be taught, or is it rather to be acquired by practice? Or is it neither to be practised nor to be learned but something that comes to men by nature or in some other way?"

According to Aristotle, possessing virtue is a matter of knowing which actions are appropriate in a range of contexts which call for an ethical response. As such it must be learnt. People are certainly not born virtuous. Socrates' inquisitor regards learning as a matter of acquiring intellectual knowledge through instruction in rules and principles. In the case of ethical knowledge, knowledge of the virtues, like generosity, intellectual understanding would be something like a grasp of a principle which one could consult in order to decide whether or not to give or share one's possessions. This is the point of the contrast with "acquired by practice."

Aristotle's answer is, in effect, that the virtues are gained by living and reflecting on a virtuous life. His conception of ethical understanding is as a species of practical knowledge: a set of dispositions to respond correctly in ethical contexts which are acquired by practice within an ethical community. One can, once the relevant dispositions are acquired, reflect on these dispositions and the principles they embody. Ethics can thus be the object of intellectual understanding, but what is primary is the practical knowledge of what to do in each situation.

It is primary for reasons to do with the inadequacy of theory as a guide to action in ethical cases. Firstly, one might doubt whether one could derive, from the limited cases to which one was exposed in one's own ethical experience, a universal principle which could cover every case. Far better to inculcate a set of dispositions which would extend in an open ended way to new situations: one could then rely on one's practical knowledge, one's inclinations about what to do, rather than having to search for a principle which covered the new case as well as the old ones. This is, I think, what Myles Burnyeat means when he says "Aristotle is pointing to our ability to internalise
from a scattered range of particular cases a general evaluative attitude which is not to be reduced to rules or precepts."\textsuperscript{24} This "general evaluative attitude" is the practical knowledge which an agent possesses which disposes her to act virtuously.

Secondly, in cases of ethical knowledge, as in other cases of intentional understanding, the distinction between conforming to and following a rule arises. For instance, the peripheral theory of the dispositions of a merciful person looks like this.

It is merciful to look after sick people

This person is sick

I must look after them

There are two ways in which one can embody this principle in one's actions. The \textbf{first} is if one has been trained to do merciful things in such a way that one does them automatically or by reflex. For instance a child who grows up in a nursing home, perhaps the daughter or son of one of the staff, may automatically care for any new inmates because she has absorbed patterns of behaviour with which she is surrounded. But for such a child, unless she is also aware of the possibilities of neglect or indifference to the patients, or the idea that mercy may extend beyond her limited exposure to the care of the sick, her actions do not exemplify the virtue of mercy. She has the basis for a practical understanding of the concept of mercy but at this early stage she is acting out of habit, and not following a rule for merciful conduct.

\textbf{Secondly}, a person may have an intellectual understanding of the merciful, or any other practical syllogism, and use it to govern their actions without thereby being merciful. For instance an anthropologist who comes to an unfamiliar community may observe a set of practices in that community and develop a theory of those practices. Since she wants to fit into the community

\textsuperscript{24}(Burnyeat p72)
she uses her intellectual understanding of the theory in order to assimilate smoothly. However she calculates the correct thing to do on each occasion according to her theory rather than responding directly to each situation as it arises. Even though, as a result, her actions conform to the ethical precepts of the community, she cannot be said to be acting ethically because she is not acting from a virtuous character. One acquires a virtuous character by expanding the range of one’s habits through education and practice until one can respond in the right way "by second nature" as Aristotle puts it. The anthropologist is not acting out of second nature but according to a strategy of assimilation.

Neither the anthropologist nor the child in the nursing home have virtuous characters, even though, for their different reasons, they may behave as if they do. Virtue is a matter of acquiring practical knowledge of what to do through training and practice so that in each new situation one responds in a characteristic way, where characteristic means not merely typical, but according to one’s character. One becomes disposed to respond in a virtuous way by being exposed to the society of virtuous people and, perhaps in the initial stages, learning in a routine and predictable way the patterns which one will later extend. The early learning and habituation are the basis for the later development of fully fledged virtue. As Aristotle puts it: "Of starting points [for the development of understanding] some are seen by induction, some by perception, some by a certain habituation...". He is in no doubt that in the case of ethical understanding habituation is the starting point.

Now some think that we are made good by nature, others by habituation, others by teaching. Nature’s part evidently does not depend on us, but as a result of some divine causes is present in those who are truly fortunate; while argument and teaching we may suspect are not powerful with all men, but the soul of the student must first have been cultivated by means of habits, for noble joy and noble hatred, like earth which is to nourish the seed... The character,
then must somehow be there already with a kinship to virtue, loving what is noble and hating what is base. (*Nichomachean Ethics* 1179b-31)

Burnyeat has a subtle discussion of this passage, relating it to Aristotle’s other ethical doctrines and his solution to the problem of moral motivation. The central theme is that from early habituation we develop the pattern of responses from which we can be further educated into a truly virtuous character. Someone who not only has the necessary patterns of response, but also some insight into them and the ability to articulate and reflect on the reasons for her responses is the completely virtuous person. However intellectual understanding is the last stage in the development of a virtue. As he says:

> From all this it follows, not only that for a long time moral development must be a less than fully rational process but also, what is less often acknowledged, that a mature morality must in large part continue to be what it originally was, a matter of responses deriving from sources other than reflective reason.  

Those other, less than fully rational, sources are the dispositions with which a person growing up in a virtuous community is inculcated. In the absence of those dispositions, one’s understanding of ethics, no matter how good the theory one has of ethical behaviour, remains purely intellectual because it is not allied to a virtuous character.

The importance of internalising ethical principles as second nature is shown by our characteristic responses to the actions of others. If we think that their generous, merciful, or other virtuous actions flow naturally from their character then our response also has an ethical component. That ethical component comes from the recognition that the agent is another person who is trying to live her life according to standards we can recognise and share. In

---

25 (Burnyeat p80)
contrast, if we think that the action is done automatically, out of pure habit, then our response does not include this component of mutual recognition or joint involvement. Nor will it if we think that the agent's action is the product of some decision calculus or purely formal mechanism. (One is not grateful for the bottle of champagne and the gold watch given by the company president on one's retirement if one knows, as one must, that every person who retires gets a gift and a speech whose value and length are computed to a formula based on years of service, level of seniority and superannuation contributions.)

The knowing how, knowing that, distinction which Ryle drew lays the basis for an account of intentional explanation which does not depend on accounting for action in terms of prior theoretical understanding. However, habits or single track dispositions do not explain the way in which such activities can be simultaneously open ended and yet turn out to accord with the rule for the activity, be it the rules of chess of grammar or of etiquette. The temptation is to revert to the theoretical account and to say that intentional agents are governed by a system of inner representations which grounds tacit inferences from past success to future suitability of behaviour.

Aristotle, however, gives us another alternative: practical knowledge of a theory of intentional activity. From a base in a series of single track dispositions or habits, the agent expands and links her capacities until, as with any fully developed ability, she is able to extend her knowhow to cover new situations. Ethical understanding is just one example of learning in general.

What begins as a set of stimulus-response patterns gets progressively articulated and refined as the children acquire from their elders an appreciation of the rich fabric of entailment relations characteristic of the theoretical framework as a whole.26

26(Churchland 1979 p29)
This quotation is not from Aristotle, but from Churchland as he attempts to explain how it is that a theory can be acquired, and can explain the actions of an agent even though the agent herself has no intellectual knowledge of the theory.

The "appreciation of the rich fabric of entailment relations" which explain an agent's actions is not the intellectual knowledge of a set of propositions and their relations, but practical knowledge of those propositions acquired against a background which allows the agent to develop her capacities through education and practice. To act in character for such an agent is just to be disposed to act, to form judgements, to speak, imagine and deliberate in ways which are consistent with each other. When Ryle says that the dispositions of an intentional agent "hang together on a common propositional hook" he is referring to this integration of a set of dispositions. Ryle's emphasis on knowhow is designed to show that such rational integration does not depend on theoretical knowledge of the propositions which describe the interrelations of an agent's dispositions.

Ryle is often presented as a reductive behaviourist, someone who thinks there is nothing more to intentional activity than patterns of behaviour. In a sense this is correct, if we allow that such things as remembering, imagining and deliberation are instances of behaviour (as we should since, on Ryle's account, they are imaginative rehearsals or extensions of skills acquired in the public realm.) However Ryle does draw a distinction between "mere" behaviour and intentional activity. He wants to hold to that distinction without explaining the intentional activity in terms of the internal application of propositional knowledge.

As a first step, Ryle needs to find a set of dispositions integrated in such a way that they can be correctly described as constituting a cognitive attitude toward some rule, standard or theory, of performance for the relevant activity. The musician learning her part is such an example. Of course her performance counts as learning rather than mere habit or rote performance because it is
acquired within a group who maintain standards for performance. The fact that
the performance can be done better or worse and is learnt and performed with
the disposition to submit to those standards of arbitration is what makes the
performance a cognitive activity. Ryle then needs to show that such intentional
activities, which give the appearance of not depending on propositional
knowledge, are not produced as a result of tacit knowledge of the theory or
some translation of it. Little is yet known about the way the information on
which intentional activity depends is stored and processed. However, Ryle has
the regress argument to show that, however the task is achieved, it is not by the
mechanical processing of symbols according to rules. This is not to deny that
where a performance depends on practical knowledge there is a peripheral
theory true of that performance. However the relation between peripheral
theory and the correct theory of the actual processing is most likely like that
between G and G*. The peripheral theory which systematises the dispositions
of an agent need not do so by mapping the structure of the internal processing
on which the activation of those dispositions depends.

It is this final point which is most important. Once the idea of a necessary
structural correlation between peripheral theory and internal processing is broken, the
way is open for an account of intentional activity in terms of practical knowledge. I
give such an account in the next chapters.

Part of that account depends on a discussion of the case of autism, in
which I reject the idea that autism can be explained as the inability to develop a
"theory of other minds." En route to my Rylean conclusion about autism I also
discuss another possible explanation for the disorder which has its basis in a
philosophical account of intentional understanding, which, while it rejects the
"theory theory" does so for reasons quite different to those advanced by Ryle.

I mention it here for two reasons. Firstly because it is an antitheoretical
account, and also because it is a philosophical account which, like Dreyfus',
relies on the phenomenology of intentional understanding.
The second reason is that argument from phenomenology is also an antisceptical account of the epistemology of other minds. Much of the literature on autism has its philosophical basis in the neo Quinean assumption that knowledge of other minds depends on the construction of an adequate theory of the causes of behaviour, and explains the autistic disability as a failure to internally represent such a theory. The phenomenological account, while it is a philosophical failure, nevertheless points us in the direction of certain aspects of intentional understanding which are not adequately captured by the theory theory.

2.5. An Alternative Non Theoretical Account. The Attitude to a Soul.

There is another, more radical, rejection of the idea that intentional explanation depends on theoretical knowledge, which is found in some of Wittgenstein's remarks and developed by, among others, Strawson, McDowell and Cook, which takes its theme from a famous remark of Wittgenstein:

My attitude toward him is an attitude towards a soul. I am not of the opinion that he has a soul...The human body is the best picture of the human soul.

This remark is famously difficult to interpret, but as a starting point we can consider the type of judgement reported by a sentence like "I saw, straight away, that she was unhappy." These sorts of reports are commonplace, especially about emotions, like depression, fear or elation. Other cases are similar. We say we can hear the anger in someone's voice, just as we can see from looking at them that they are angry. If one takes these sorts of statements as true simpliciter, then the notion that intentional explanations are based in perception of other's mental states follows directly.

Another thing which follows is that in our relations with others we do not consciously infer the nature of their mental states from their expression or
bodily movement (i.e. we do not form opinions about the nature of their soul from the state of their body. On this point see the discussion at 5.6.)

Wittgenstein is clearly offering an argument against Cartesian scepticism here: he is rejecting the view that the body is an inanimate physical substance which moves and lives according to the dictates of the thinking mind. The exact nature of this anti Cartesian argument is contested. Some have presented him as offering a version of logical behaviourism, similar to the Rylean position I discussed earlier.27

However the Rylean position is concerned to refute the sceptic by showing that we can have secure practical knowledge of other minds conceived of as sets of dispositions. To that extent it accepts the sceptical problem "How can we know what others are thinking, or if they have minds at all?" on its own terms and offers a reply which reconceives of knowledge of minds in the practical way.

However, there is another way of understanding Wittgenstein's purpose in these passages. That is to treat him, not as offering a solution to the sceptical problem but denying, at least for our ordinary way of speaking, thinking and responding to one another, that there can be such a problem. The idea that there are distinct bodies and minds and an epistemological problem of linking our knowledge of minds (which can only be observed in introspection and are therefore unknowable to third persons) to our public observations is a mistake. The source of the mistake is, of course, the idea that we must construct our knowledge of other minds in the same way we construct the rest of our empirical knowledge: by theory based on observation.

As an alternative, Wittgenstein offers the view that, in our transactions with other persons, we are pretheoretically committed to their being minded. We automatically treat other persons as thinking and feeling in the same way as ourselves, and automatically treat their words, expressions and gestures as expressive of states of mind. The sceptical position invites us to assume that

27 see, for example, (Chihara p283)
there are, or can be, bodies and mental states which can come apart, so that the question is always open whether a particular body is animated by mental states. Wittgenstein is pointing out that when we are relating to other persons this question simply does not arise:

Being sure that someone else is in pain, doubting whether he is, and so on, are so many natural, instinctive kinds of behaviour towards other human beings.

As a piece of descriptive metaphysics, the "metaphysics which is content to describe the actual structure of our thought about the world" 28 this is correct. Our ordinary notions of the world and other people who inhabit it, our everyday practices and responses treat the person, not the observable body, or the unobservable mind behind the body as primitive. 29 Consequently it makes sense to say that we see emotions, in the sense that when we see someone crying we see that they are unhappy or in pain.

Those who follow Wittgenstein on this point often develop an argument that descriptive, as opposed to revisionary metaphysics, are primary in matters connected with ethics, persons and language. In particular they regard any attempt to reduce the meaningful human activities to facts described impersonally in the language of science as misconceived. 30 They block any attempt to adopt the standpoint from which the sceptical question arises. Thus they also rule out the neo Quinean position, which solves the sceptical problem by regarding mental states as the theoretical postulates of a well confirmed empirical theory, as an illegitimate response to an illegitimate question. 31

28 The terms descriptive and revisionary metaphysics are Strawson's.
29 To this extent competing theories of personal identity which attempt to establish that the extension of our everyday concept of personhood is actually the physical body, or a coherent psychology or a four dimensional space time worm are revisionary.
30 McDowell being a classic example. See (McDowell 1981)
31 McDowell, for example, regards empirical scientific explanation of linguistic behaviour, as committed to either behaviourism or "Psychologism" an account of the brain processes which cause the behaviour. Either approach has nothing to say on the question of how human beings read meaning into each other's activities, according to McDowell, because meaning is the result of immersion in activity which depends on prior recognitional capacities "if we eliminate dependence on prior involvement in forms of life, then we eliminate the very possibility of understanding" (p237). McDowell cannot make sense of the suggestion, seriously advanced by
One way to interpret this Wittgensteinian position is to regard the neo-Quinean, who treats intentional explanation as proto science, as engaging in revisionary metaphysics "concerned to produce a better structure" of our everyday thought about the world. The dialectic between the two positions is then clear. The Wittgensteinian maintains that intentional explanation is based in an everyday practice for which the nature of other minds and the possibility of their non-existence, non-embodiment or realisation in another physical structure does not arise. Persons, *qua* human beings with minds, are primitive. Hence the remark:

Only of a living human being and what resembles (behaves like) a living human being can one say: it has sensations; it sees; is blind; is deaf is conscious or unconscious.

The neo-Quinean, on the other hand, while he may agree with, or be agnostic about, the Wittgensteinian descriptive metaphysics, says that we should revise our understanding of mental states by putting them on the same metaphysical basis as other objects of scientific investigation. Thus he asks us to reconceive of intentional explanation as protoscientific psychology. From this perspective, mental states are not essentially linked to human beings, bodies and their intersubjective responses because we investigate mental states, not in terms of the role they play in our lives, but *sub specie aeternitatus*.. This is why the Cartesian perspective, which generates scepticism about other minds leads so inexorably to the project of cognitive science: scepticism requires us to justify our practice of attributing mental states to each other. This challenge most naturally invites a response in terms of satisfaction of an empirically adequate

---

Dummett, that a good theory of meaning would be one which a Martian, or someone who does not share our form of life, or perceptual capacities, could use to understand our words. The best exposition of this aspect of Wittgenstein's thought, including his dismissal of attempts to rebut the sceptic by constructing a reductive behaviouristic or Cartesian theory is by (Cook 1969). Anscombe echoes some of the same themes in her "The First Person" where she says that "I" apparently refers either to a Cartesian ego or to an organismic body both of which are inadequate to our concept of personhood." (Anscombe 1957) Strawson put the point more elegantly when he said "the concept of a person is not to be analysed as that of an animated body or embodied anima."
theory. Hence the Wittgensteinian hostility to "scientism" in the philosophy of mind, the attempt to understand mental states in terms of theories of cognitive processing.

In the next chapter I consider some explanations of the acquisition by young children of the concepts of intentional psychology which are explicitly committed to the scientific view Wittgenstein rejects. While I do not endorse Wittgenstein's philosophical objection to the project, it turns out that the theory of cognitive processing I offer as an alternative to the mainstream "theory theory" explains many of the intuitions on which the Wittgensteinian relies.
Chapter Three: The Child’s Theory of Mind

Precis

The chapter examines the empirical evidence for the hypothesis that children’s acquisition of the concepts of intentional psychology depends on the development, through a series of precursor stages, of a modular capacity for the representation of psychological theories: The Theory of Mind Module (ToMM).

3.1. The Child’s Acquisition of Language. Language ability supported by a modular capacity for the representation of grammatical principles.


3.3. The Child’s Theory of Mind. The principle extended to explain the child’s acquisition of psychological concepts.

3.3.1. CTM and ToMM: Modularity. The different constraints imposed by theories which explain behaviour as a result of the activation of an innate modular capacity for representation of psychological concepts, or the operation of a general cognitive ability applied to the psychological domain. The concept of an M-representation explained: a mental state which represents an agent, a proposition and the agent’s attitude to that proposition.

3.3.2. The False Belief Test. The experimental evidence on which CTM and ToMM rely; the ”False Belief Test” designed to test for the understanding of psychological concepts, construed in the manner suggested by Churchland at 1.4. Once again, however, the theory is ambiguous between practical and intellectual understanding of psychological concepts.

3.3.3. Further Evidence for ToMM? Further experiments designed to disambiguate the False Belief results. The conclusion is that the experiments test for ability with psychological concepts, not a general facility with the concept of representation.

3.4. Theory of Mind. Constructed or Innate? The innateness hypothesis reconsidered in the light of the preliminary evidence. While the evidence tends to support the view that the child has some innate abilities to interpret behaviour, it does not support the view that all those abilities
depend on the functioning of a module devoted to the representation of psychological concepts.

3.4.1. Early Stages in the Acquisition of a Theory of Mind? Psychological concepts are typically acquired in a sequence. The ToMM hypothesis is that sequence can be best explained in terms of the sequential development of a module which represents psychological concepts.

3.4.2. The Understanding of Opacity. The M-representation hypothesis is shown to be a product of the need to account for the opacity of mental states invoked in behavioural explanation. Acquisition of specific concepts.

3.4.3. Understanding the Concept of Pretence. Pretence considered as a precursor.

3.5. Precursor States. The concept of a precursor explained. Two possible interpretations. The M-representation hypothesis typically goes with the less plausible view of precursors. The mistake involved is in assuming a structural isomorphism between the peripheral theory which explains children’s behaviour and the nature of the representations on which that behaviour depends: i.e. explaining practical knowledge in terms of explicitly represented intellectual knowledge.

3.6. Conclusion Precursor to Later Chapters

This chapter examines a claim that children progressively develop a series of hypotheses about the causes of others’ behaviour, culminating in the theoretical conception of intentional psychology outlined at 1.4. In this chapter I claim that the evidence for this claim supports the conclusion that children acquire practical knowledge of the truths of intentional psychology, rather than intellectual knowledge systematised as a causal theory of action. In particular, the evidence does not support the conclusion that the child who successfully interprets and explains behaviour does so by tacitly representing a hypothesis about the nature of the brain states which cause behaviour. Thus, as we would expect if the last chapter is correct, the fact that children acquire an understanding of the truths of intentional psychology is not yet evidence for any particular hypothesis about the way those truths are represented. It may be the case that intentional understanding depends on the function of a "theory of
mind" module of the brain, but that is an empirical hypothesis not established by noting that the surface behaviour of children is consistent with that possibility. In chapter five I develop an alternative hypothesis about the nature of the representations which sustain practical knowledge of intentional psychology and argue that it does just as well on the standard cases and is a superior explanation of the deviant cases, in particular autism, which are cited as evidence for the "theory of mind" hypothesis.

Human infants are born with very few cognitive or physical abilities. Yet by the time a child is four it has, according to contemporary developmental psychology, acquired the ability to explain others in terms of their intentional mental states. This very rapid development of a sophisticated psychological ability parallels (and in fact takes place at roughly the same time as) the acquisition of language. Interestingly, the paradigm which presently structures research into infant psychology explains the acquisition of intentional concepts in terms of the maturation of a modular capacity of the brain.

Developmental psychologists tend to treat tacit knowledge as internalised intellectual knowledge, which commits them to very specific accounts of the way tacit knowledge must be realised. All the theorists I discuss converge on the idea that our understanding of other people depends on mental representations with a certain structure: M-representations. M-representations represent an agent, an attitude, such as belief or hope, and a proposition, or content of that attitude. Thus the content of the attitude may be true or false, without affecting the truth of the M-representation.

To understand the attractions of this model and in order to discuss the evidence which must be explained by my alternative account we need to see why it was first postulated.

---

1 It is worth noting here that in some cases surface behaviour may be evidence of the nature of underlying mental structures. Chomskians, for example, claim plausibly that the nature of spoken language is such that it can only be generated by a mental mechanism which relies on specific grammatical principles. The claim splits into two two parts: that every language spoken conforms to certain grammatical principles and that languages which do not conform to those principles cannot be understood by humans. The same claim might be true of other mental abilities. The ability to make certain visual discriminations and not others implies that the visual system of an organism only represents certain wavelengths.
A grammar for a language can, as I said at 2.2.2., be understood in two ways. Firstly, as a peripheral theory which systematises rules for correct or meaningful usage. In this sense a grammar is independent of the psychological processes which underlie the correct production of language: i.e. it does not attempt to explain how the practical knowledge of competent language users is realised.

However, a grammar may also be used to describe and explain those psychological processes themselves. Such grammars attempt to explain, not just the surface regularities in language use, but the structures which must be represented in the brain of the language user in order to produce grammatical sentences. An example is the way generative grammars explain the stages through which children pass on the way to mastering their native language. An instance is the progressive mastery of the past tense in English. In the first stage children are proficient with the past tense of verbs they have actually heard or used themselves. So a child will correctly say "I ran to the shops." At this stage, although the child's usage is correct it is based on reproduction of what she has heard, rather than produced according to a grammatical rule. However, shortly after, the child starts to make characteristic mistakes in the use of the past tense, even for verbs which she used to decline correctly. So the child who used to say "I ran to the shops" will now say "I runned to the shops." At this stage the child is internally representing and applying a general rule for the past tense (ad the "ed" suffix for actions which happened in the past). This rule extends to unfamiliar cases, so that all past actions are described by verbs using the "ed" suffix, whether or not the child has heard the appropriate verb declined before. So the child, although she makes mistakes, makes them...
systematically across a range of familiar and unfamiliar cases via the application of a simple rule. Still later the child masters the declension of irregular verbs, and will revert to saying "I ran." At this last stage the child is producing her language according to a still more sophisticated rule which distinguishes different grammatical structures for different verbs.

If we regard a grammar as a theory of correct language use then it is appropriate to say that the child has progressively acquired the correct theory of the use of the past tense in English.

One consideration in favour of explaining the child's language ability in this way is that the stages exhibit a typical pattern of theory acquisition. Theories postulate unifying principles, laws or rules, to explain a range of data. Consequently, as new or inconsistent data is encountered, the explanatory principles which make up the theory are extended or modified. If sufficient inconsistent evidence is encountered, the theory must be completely revised or discarded and a new theory installed.

The test of theory change is whether the same set of evidence is interpreted differently, or generates a different set of predictions, via appeal to different explanatory principles. (Recall 1.3 in which a defining feature of a theory was its systematisation of a domain in terms of a set of underlying principles). A classic instance in the philosophy of science is the change in astronomical theory from the Ptolemaic, heliocentric, theory of circular planetary motion to the Keplerian theory of geocentric, elliptical, orbits. Both these theories systematise the same observational evidence, the movement of the sun and the planets, in terms of different fundamental principles. However, where Ptolemy placed the earth at the centre of the solar system, Kepler placed the sun in the middle.

The Ptolemaic theory was ultimately replaced because of its inability to explain the trajectories of the other planets in relation to the earth and the sun. However it was not discarded straight away. Initially these inconsistencies with the heliocentric theory were reconciled by modifying the initial theory and
adding some auxiliary hypotheses. An example is the heliogeocentric theory of Tycho Brahe, in which the planets orbited around the sun, which in turn orbited the earth.

These theoretical transitions show how theories are generated and revised in order to cover an increasing range of evidence. Part of that process of theory revision is the postulation of auxiliary hypotheses, such as that of Brahe, to extend the range of the initial theory.

The same pattern is evident in the child's acquisition of language. As she tries to understand the range of cases to which she has been exposed she generates explanatory hypotheses. The first of these in the language case, might be the simple theory, "all verbs describing action in the past tense end in 'ed'". This rule is inadequate to the complexity of English and is eventually replaced by a more complex theory of the past tense. In the interim period the child's speech exhibits all the characteristics of the transitional phase in scientific theory: some unfamiliar verbs receiving the "ed" treatment, others tentatively declined as irregular verbs on the basis of phonetic similarity to familiar cases. The simple theory is maintained via the addition of auxiliary hypotheses to cover those, irregular, cases which appear as counter evidence.

Explaining cognitive abilities as the mastery of a theory of the relevant cognitive domain unifies the study of cognition. As Gopnik puts it

scientific theory and conceptual change in childhood are both the product of human minds trying to understand the world around them. From this point of view both cognitive developmentalists and philosophers of science face a similar question...how to characterise mechanisms for learning about the world. Two of the clearest and most impressive instances we have of such mechanisms are the acquisitions of children and the achievements of science... The hypothesis of the theory theory is that there are deep similarities between the underlying cognitive mechanisms involved... But the similarities are better captured by thinking of scientists as big children rather than thinking of children as little scientists. The
progress of science, we believe, reflects certain fundamental processes of conceptual change that are first seen in very young children.³

As I said earlier, claims about theoretical knowledge can be taken in two ways: as a claim that people have practical knowledge of the theory, subtended by a range of dispositional representations, or that they explicitly represent and take attitudes towards its propositional content. Scientists obviously do the latter. CTM theorists have argued that children learn to understand other people by progressively adopting and discarding internally represented theories about the causes of their behaviour.

3.2. The Chimpanzee Theory of Mind

In the previous section I mentioned that one reason for regarding explanations as depending on theories is a change in the predictions or interpretations that an agent makes. Where such a change is justified by reference to different explanatory principles then we can justifiably say that one theory has succeeded another. Another, related, reason is the use a prediction or interpretation makes of entities whose properties are not directly observed, but inferred from observational evidence. As we saw in the first chapter, the relevant theoretical entities in the case of intentional explanations are states of the agent's mind, whose properties are inferred from the agent's behaviour.

Chimpanzee, and Child, theorists of mind interpret their evidence as showing that the child makes these sorts of inference, but they do not give an account of inference itself. Inferences are transitions between beliefs such as the drawing of conclusions from premises or the revision of beliefs in the face of evidence according to some epistemic principle. In attributing someone with the ability to make particular inferences the same practical/theoretical distinction arises as for any attribution of belief. An agent may explicitly represent a logical principle or epistemic heuristic and consult it in forming or

³(Gopnik In Press p4)
revising their beliefs, as scientists do when interpreting evidence or logicians constructing proofs. Alternatively, an agent may exhibit behaviour which can be described by attributing attitudes to propositions which stand in particular inferential relations, without it being the case that the agent controls its behaviour by any explicit act of inference. So there are practical and intellectual inferences just as there are practical and intellectual beliefs. And just as practical knowledge does not always imply explicitly represented intellectual knowledge, practical inference does not straightforwardly imply the presence of an internal inference engine which operates by explicit representation of inferential principles. Chimpanzee, and Child, theorists of mind generally do not draw this crucial distinction. The result is that they attribute implicit theoretical knowledge on the basis of behavioural evidence without explaining how that knowledge is represented.

Chimpanzee society is a volatile phallocracy in which males constantly compete for top (alpha) ranking. The relative ranking of males determines mating precedence, access to food and the protection of other members of the troupe. However, competition is seldom between pairs of chimpanzees but is usually complicated by the intervention of a third, allied to one of the competitors. Thus the success of a chimpanzee in competition depends on its ability to recruit allies. These alliances are stable over long periods but as older chimpanzees die, new ones reach adulthood, or the physical environment changes, the alliances are reformed.

As well as the formation of alliances, chimpanzee politics is characterised by deceptive behaviour. Chimpanzees will try and deceive others about the location of food or their actions. So a chimpanzee who has discovered a source of food will walk away when another approaches and only return to eat the food source when the others cannot see it. Even more sophisticated behaviour

---

4 Another way to put the distinction would be to say that in the case of intellectual inference the relevant brain processing involves semantically interpretable representations of propositions, their elements and rules for their interaction. In the case of practical inference the brain processing merely gives rise to behaviour interpretable as the derivation of conclusions from premises or evidence.
has been reported in which competitors leave the scene and conceal themselves, watching until the first chimpanzee reveals source of the food, whence they return and take it. Sexual behaviour is also characterised by acts of deception:

often a low ranking male will sit with his upper arm resting on his knee and his hand hanging down in front of him so that a female in front of him can see his erect penis, but apes on the side cannot see it. This inconspicuous form of concealment occurs together with quick sideways glances at dominant males.\(^5\)

Bearing in mind the complex and competitive nature of chimpanzee society and the importance to survival of alliance formation and tactical deception, a chimpanzee who was able to understand and predict behaviour would have an advantage over its competitors. As Humphreys puts it, there is an evolutionary advantage for the chimpanzee who can become the best chimpanzee psychologist.

Chimpanzees, it seems, are able to make these sorts of predictions. Of course this ability is not yet evidence of a chimpanzee theory of mind in the sense advocated by the neo Quinean and CTM theorist. The chimpanzee’s abilities are only evidence of a theory of primate psychology if the chimpanzee interprets the behaviour of its fellows and controls its own behaviour by inferring the intentions of its fellows from their behaviour and making further inferences about their likely behaviour given those intentions.

However, the evidence I have cited so far shows only that chimpanzees are able to manipulate the behaviour of their fellows. A skill attributable to their ability to recognise typical correlations between what a chimpanzee sees, or does not see, and what it does. The theory of mind hypothesis requires that the chimpanzee form a concept of intention as a state which intervenes between perception and action and employs that concept in dealing with its fellows.

\(^5\) Described by DeWaal, quoted by Whiten.(Whiten 1993 p 377)
Evidence for that additional step was first suggested in a series of experiments by Premack and Woodruff with Sarah the chimpanzee in 1978. Sarah was shown videotapes of people trying to solve problems, lighting a fire, reaching for a banana etc. Then she was given a further series of photos, only one of which showed the solution to the problem. Sarah unerringly chose the correct photo. As Premack and Dasser put it, Sarah's ability is best explained on the assumption that she conceived of the video sequences as attempts to solve a problem.

Once the jumping individual is perceived as having the intention of getting the bananas, as wanting the bananas and trying to get them, the sequence becomes a problem, and its becomes sensible to choose solutions. But not until the sequence is interpreted in this general manner does the consistent choice of solutions make any sense. 6

Of course this is a bold assumption. The evidence can also be taken to favour a more conservative conclusion: that the chimpanzee actually recognises patterns of behaviour and makes predictions about the way the sequence should continue, without positing any intervening intentions. 7 (One way to test for this possibility would be to scramble the sequence by including irrelevant actions in the sequence and then including photos of those actions among possible solutions. If the chimpanzee still successfully chose the correct solution this might support the hypothesis that it was using something like a concept of actor's intentions to sort the irrelevant from the relevant action. This experiment has been carried out with children, who successfully completed the sequence, but not with chimpanzees.8)

---

6 (Premack 1991)
7 If one is a behaviourist then the recognition and projection of behavioural patterns is all there is to the attribution of intentional states.
8 (Premack 1991)
Chapter 3  

The Child’s Theory of Mind

So the notion that chimpanzees explain others’ behaviour in terms of a link between perception, intention/mental state and behaviour has not been conclusively vindicated by experiment. Support for the contention rests on the way it would explain and integrate a range of abilities (deception, alliance recruitment, sequencing, problem solving) the chimpanzees possess. If we regard the chimpanzee, for example, as possessing a concept of hunger, implicated in others’ food seeking behaviour then the chimpanzee’s responses in a wide variety of competitive food seeking and novel experimental situations are explained economically. Without such a hypothesis each separate instance of banana seeking, fridge searching, food concealment etc. needs to be explained independently. If, however, we regard the chimpanzee as positing a state of hunger, implicating a food-seeking intention, which explains its companions’ behaviour, then we have a unifying explanation of a range of chimpanzee behaviour, natural and experimental.

If this hypothesis is accepted, then we attribute the chimpanzee with mastery of the concept of some mental state linked to perceptual experience. To say the chimp knows that another chimp is hungry or wants food is not, however, to make any claim about the nature of that knowledge. The most economical attribution is to attribute to the chimpanzee practical mastery of the concept of such a mental state; i.e. to say that it has abilities, however realised, which amount to knowledge of the concept of a state which plays a causal role in the production of action.9

Some theorists, however, have interpreted experimental results to yield a stronger conclusion. For example, chimpanzees have been shown a scene in which a banana is hidden in the presence of two human observers, one of whom is blindfolded. The chimpanzees typically enlist the help of the non-blindfolded observer in their subsequent quest for the banana. If a banana is hidden in the presence of a blindfolded person the first thing the chimpanzee

---

9 Whiten argues for a similarly modest interpretation of claims of a chimpanzee “theory of mind.” (Whiten 1993pp384ff)
will do when it commences its quest is to remove the blindfold from its human assistant.

Similar abilities in gorillas have led Gomez to conclude "she seemed to understand that in subjects perceiving is causally related to acting. And here is where the mind appears, since the coordination between perception and acting is carried out by the mind." Gomez call this an "implicit" theory of mind but the term is crucially ambiguous between the practical and intellectual understanding of theoretical knowledge. Actually Gomez seems to mean something like practical knowledge of a theory of mind.10

However, even if the gorilla or chimpanzee acts on the basis of practical knowledge that certain actions are (typically) produced as a result of perception, such actions do not imply any particular account of the mental representations on which they depend. In particular, as Gomez, Whiten and Heyes realise, the chimpanzees' behaviour does not automatically support the hypothesis that they control their behaviour by explicitly representing a theory of intentional psychology which postulates internal states with specific properties. With these theorists, I believe that in the absence of any conclusive evidence we should remain neutral on competing hypotheses about the way the chimpanzee controls its behaviour.

This is not to deny that chimpanzees can have practical knowledge of a theory which uses mentalistic notions like "pretend" or "deceive". One reason why chimpanzees attract this type of mentalistic attribution is that their behaviour is flexible and adaptive. It is not just the reiteration of patterns of stimulus and response, but the extension from previous experience of abilities which may have been required in limited, even stereotypical situations, to cover new circumstances. The experiments and observations of chimpanzees show that chimpanzees acquire a set of multiracked dispositions, not just a set of stereotypical responses. Chimpanzees understand from their encounters with the dominant male, not just that they will be bitten if they repeat the

10 (Gomez 1991)
behaviour which provoked the initial biting, but that there are a range of behaviours which will provoke not merely biting, but other manifestations of dominance. This is just one instance of the chimpanzees' ability to solve problems which require them to extend from their practical knowledge base. However, as Ryle argued, this type of ability does not necessitate the stronger hypothesis that the chimpanzees' knowledge depends on the explicit representation of a theory of "other minds".  

11 Sterelny and Celia Hayes are others who argue for the modest interpretation of the chimpanzees' abilities. Sterelny explicitly concludes: "our best guess is that the selective regime is one in which chimps have acquired a sensitivity to, and the capacity to respond to, complex dispositions of their fellows" (Sterelny 1995)
When CTM theorists refer to the child’s "theory" of mind they claim that by the time the normal child is three and a half to four years old she ascribes to others and herself, mental states with distinctive properties. The four year old has, according to CTM, a concept of intentional states as internally represented attitudes to propositions. So CTM satisfies the constraint of theoreticity in being an articulated explanatory hypothesis about human behaviour: namely that humans act on the basis of their representations of the world and of the objects of their desires, representations which have a propositional form. Gopnik is quite explicit about the nature of the claim made by CTM.

The developmental evidence suggests that children construct a coherent, abstract account of the mind which enables them to explain and predict psychological phenomena. Although this theory is implicit rather than explicit, this kind of cognitive structure appears to share many features with a scientific theory. Children's theories of mind postulate unobserved entities (beliefs and desires) and laws connecting them. Their theories allow prediction and they change (eventually) as a result of falsifying evidence. Moreover the child's theory of mind is equally applicable to the self and others.\footnote{Gopnik 1993 p10. For a detailed discussion see also (Gopnik 1992)}

An argument in favour of CTM is that the four year old child's understanding of behaviour develops through a series of stages. The four year old child, for example, employs a different set of concepts in understanding behaviour to the three year old, the eighteen month old and the newborn. CTM explains the succession of these sets of concepts as a series of theoretical developments. The four year old's theory of mind is the last in a series of theories the child adopts to interpret and explain the evidence presented to her by the behaviour of other people.
According to CTM, the child is doing the same sort of thing as a geneticist who infers the nature of a gene from the colour of the flowers on a tree, or a scientist who infers the nature of gravity from the behaviour of the planets. Obviously the mature scientist grounds her inferences in a theory which is much more cognitively sophisticated, articulating the relationship between theoretical variables mathematically for example, but there is no essential difference between the child and the scientist, just a difference in the sophistication of their theories and their subject matter.

There is another version of CTM, ToMM, (short for The Theory of Mind Module) which argues that the findings of Gopnik and others show that the child’s understanding of others is based on an innate modular ability to form and employ intentional concepts. Both versions converge on the idea that the child uses M-representations in order to explain other people. CTM proposes that the child develops a concept of other minds, based on its ability to form M-representations, by theorising about the behaviour of others and discarding a succession of empirically inadequate predecessor theories. M-representation by infants stands to the babies’ understanding of other minds as Newtonian to Ptolemaic cosmology. It is a better theory about the reason for certain observable regularities.

3.3.1. The Difference between CTM and ToMM: Modularity.

The essential contrast with CTM is that ToMM treats the acquisition of concepts depending on M-representation as the result of the innately determined development of a modular capacity. A key element of modularity, and one which is often thought to insulate cognitive activities from explanation in terms of modularity, is informational encapsulation. That is to say that the information specific to one cognitive module is not affected by information available to another. This seems obvious in the case of perceptual modules: the

13(Baron Cohen 1985; Leslie 1991a)
information processed by the auditory module is not the same as that processed by the visual module. At the cognitive level, the level at which inferences are drawn from perceptual evidence, the thesis of modularity is more difficult to sustain. After all, cognition is the process of integrating information from a wide variety of sources and generating explanations. Chemical theory is not cognitively insulated from physical theory or genetics, and the myths of Homer are equally available to Freud and Quine to use as evidence in support of their respective theories. While this is true there is evidence that some information processing abilities are at least partially modular. Consider language learning by children. If the Chomskians are correct, language ability depends on a specific module dedicated to the learning of language. Of course the information about syntax and phonology presented to the infant language user is also available to be processed at a more general cognitive level. Adults who learn a second language often take this more explicitly theoretical approach to the task. However, the difference in the result is often noted and is put down to the fact that at a certain stage in life children devote a dedicated cognitive capacity to language learning which shuts down after adolescence. Adults who wish to learn a second language have to rely on the more general, all-purpose, cognitive ability.

The hypothesis of the ToMM theorists I discuss is that understanding what other people are thinking is not the result of applying a general purpose inference engine to the domain of human behaviour but of activating a mental capacity specifically devoted to the task. It is thus similar in one respect to the Chomskian story about language learning. However, and this is a point to which I will return in the conclusion to this chapter, modularity, in the sense discussed above, is not established by showing that there is a special purpose learning device. Crudely, modularity is the thesis that the device is sensitive to specific types of information and insensitive to others. So, if the purpose, understanding other minds, can be realised by processing information which comes from different sources in widely different forms, then the fact that it is
used for the purpose of understanding other minds does not establish that the information processing device is modular. To show that one needs to show that the device is *informationally encapsulated*. ToMM theorists succeed only in showing that there appears to be a specific purpose learning device. Those I discuss do not establish that device works by processing information which is encapsulated.

The difference between CTM and ToMM is the explanation each will give of the stages preceding intentional understanding. Gopnik will explain them as a series of hypotheses about the social environment generated by the all purpose inference engine. ToMM can agree that there is a succession of theoretical models employed by children, but claims that the cognitive mechanism which subserves them is dedicated solely to the formation of hypotheses about the mental states which explain behaviour. Some ToMM theorists also regard the sequence of theoretical models employed on the way to mature ToMM as determined by genetic information rather than arising from the confrontation with counter evidence and new information. So ToMM comes in two forms, one of which adds the claim that the child is born with some less sophisticated abilities which are transformed, in a sequence predetermined by genetic information, into the ability to form M-representations.

So, while the issues of modularity and innateness are logically separable there is a substantial body of psychological literature in support of the view that the ability to understand other people's behaviour, like the ability to understand language, depends on a cognitive capacity whose development is genetically determined.

As is well known, there are empirical arguments, deriving principally from the ability of language users to produce and understand an infinite number of sentences from their exposure to a finite number of degraded and incomplete linguistic expressions, for the innateness and modularity of the

---

14 The classic exposition is (Fodor1983)
cognitive system which subtends language acquisition. ToMM is a similarly empirical hypothesis. It gains part of its plausibility from some considerations about the nature of intentionality and is backed up by interpretations of evidence which suggest a determinate developmental trajectory for the ability to entertain a concept of belief.

According to ToMM, what distinguishes four year olds' understanding of other people from three year olds' is that the four year old can mentally represent others as having attitudes to propositions in order to explain their behaviour. So a four year old who sees another person see a toy placed in a basket, forms the belief that the other person has a belief that the toy is in the basket. The four year old thus has a belief about a belief — a metabelief. The four year old who has the belief that "Sally believes there is a toy in the basket" has a belief whose contents are representations of Sally, Sally's mental representation of the toy in the basket, and Sally's attitude of belief to that proposition represented.

It is important to note that the four year old represents Sally as believing the propositions she represents. The concept of belief as an attitude to a proposition is the crucial aspect of ToMM because it signifies understanding that mental states and the world may diverge. As we saw in chapter one, in order to explain why people act the way they do, we must attribute to them states whose content reflects the subject's way of conceiving of the world, rather than states which directly mirror the way the world is.

15Unlike a priori arguments like those of Fodor and Evans for the systematicity and productivity of thought.

16 Some psychologists and philosophers have argued that in understanding the relation between perception, belief and action this epistemological issue has been exaggerated. They argue that we should focus on the basic case, veridical perception leading to true belief about the world, when we theorise about perception and belief. The epistemology of philosophers takes too seriously the prospect of divergence, when what needs to be explained is the basic case in which animals and humans successfully negotiate their environment on the basis of their veridical perceptions. These philosophers and psychologists think that we should focus on the basic case in constructing accounts of the relation between perception, belief and action. However it is not so easy to discard epistemological questions where the ability to be explained is the ability to understand other people by attributing to them mental states. The false belief experiment was designed to test a suggestion of Dennett's made in response to claims of a chimpanzee theory of mind. (Dennett 1978) Dennett suggested that Premack's results did not establish unambiguously whether the chimpanzee was employing a concept of belief similar to that employed by humans in intentional explanation, or, instead, doing some sophisticated
So the ability to form M-representations is crucial, because only M-representations represent an agent as having an attitude to a proposition which can be true or false. I explain John's behaviour in carrying an umbrella out the door by saying "John believes that it's raining." It does not matter to the truth of this attribution whether or not it is raining because the attribution tracks the way John is thinking of the world, rather than the way the world actually is. The ToMM theory is a specific theory about, firstly, the nature of the mental processes on which such attributions rely, the entertaining of M-representations, and, secondly, the way those processes develop in infancy. The crucial experimental support for the claim that four year olds rely on M-representations in interpreting behaviour is the "false belief" test.17

3.3.2. The False Belief Test

In the false belief test a child watches this scenario, or a version of it. Sally sits at a table. On the table are a basket, containing a toy, and a box. Sally leaves the room. While she is out of the room the experimenter moves the toy from the basket to the box. Sally then reenters the room. The child is then asked "where will Sally look for the toy?" Four year olds get this question right, three year olds get it wrong.

Baron-Cohen interprets the test this way "As can be seen, the test involves appreciating that since Sally was absent when her toy was moved from its original location she won't know it was moved and therefore must still believe it is in its original location."17

The false belief test, and variations thereof conducted on normal children and those with developmental disabilities, such as Down's syndrome and prediction on the basis of patterns of behaviour. He suggested that an experiment was needed which would show, not only that the chimpanzee could link perception to action, but which could establish that the chimpanzee made that link using the concept of an intervening state which would play the same behavioural role irrespective of whether or not it was veridical.17

17(Baron-Cohen 1993bp60)
autism, indicate that normal four year olds understand that people act on the basis of their beliefs about the world, not the way the world actually is. Consequently, when they see, or are told, about a situation in which an agent has beliefs or desires which diverge from the situation in the actual world, four year olds predict that the agent will act according to her beliefs, not the actual state of the world. As Wellman puts it:

by four years, normal children's understanding of the mental mediation of experience is so robust that they understand even such a seemingly veridical experience as perception itself (for example my perception that that is a flower) is a mental representation akin to belief rather than fact. That is, by four years children understand that persons are prey, not only to false beliefs, but to false perceptions.\(^\text{18}\)

However three year olds when asked "where will Sally look for the toy?" reply, or point to, "the box." The explanation offered by CTM theorists is that the three year old explains actions in terms of the environment as she perceives it, rather than the environment as represented from the perspective of the agent, as do four year olds and adults.

According to ToMM, the three year old does regard subjects as having states of mind, and may even conceive of these as representational, but these states of mind are conceived of as directly caused by the environment and "representing aspects of the world in an accurate, literal and faithful way." \(^\text{19}\) The three year old has not yet conceived of mental states which may diverge in content from the way the world is presented to her.

The false belief test is a very robust experimental result and the age at which children normally pass it has been conclusively established at around three and a half. Perhaps, however, representational theorists of mind have overinterpreted the result and there is a more conservative characterisation

\(^\text{18}\)(Wellman 1993)  
\(^\text{19}\)(Leslie 1987p414)
available which does not postulate the ability of children to represent the
beliefs of others?

3.3.3. Further experimental evidence for ToMM?

The conservative interpretation of non human primate behaviour was that
primates interpret others' behaviour, not by projecting representations into
their heads, but by making inferences about the correlation of typical
behaviours with perception of the state of the world. The chimpanzee Sarah
should not have been interpreted by Premack as solving a problem by inferring
the content of the subject's desires, but as completing a typical behavioural
sequence. Premack designed an experiment, eventually conducted with
children not chimpanzees, to test for the possibility that subjects were solving
the problem by recognising behavioural sequence rather than attributing
intentions. In these experiments actions irrelevant to solving the problem
(hitching trousers etc.) were included to scramble the sequence. However
children correctly filtered out these irrelevancies. If they were merely acting on
the basis of "what comes next" then some of their predictions should have
included irrelevant actions. Premack concluded:

one does not predict the behaviour of the other one simply on the basis of the
temporal order of his behavioural sequence. Instead, one attributes intentions to the
other one. Why? the principal gain may be a reduction in one's perceptual and
mnemonic burden. Having attributed an intention, one need no longer perceive and
seek to remember every thing the individual does. Instead one can ignore irrelevant
acts — acts that do not satisfy an intention — and attend solely to those that are
relevant. Notice that this presupposes that one can discriminate between acts which
do and do not satisfy an intention. 20

20(Premack 1991p264)
The ToMM interpretation of the false belief test suggests that there should be no difference in three year old predictions/completions of behavioural sequences implicating the beliefs and those which do not. Four year olds, however, who have acquired the concept of belief, should perform better on sequences involving belief.

To my knowledge an experiment designed to test the relative abilities of three and four year olds to disambiguate sequences involving beliefs, rather than intentions or desires had not been done. This may be due to the difficulty of designing a sequencing experiment in which the only independent variable between two sequences is that one involves beliefs and the other does not.

However, there are recognition and memory tasks in which the independent variable is a belief. The "False Photograph" test modifies the "False Belief" test in this way. To pass the false belief test the child needs to understand that Sally is acting on a belief which has become false because the situation represented by the belief has changed. In the false photograph test children are shown a photograph or a map which goes out of date in the same way as Sally's belief in the false belief test. That is, they see Sally take a picture of the scene. She puts the picture down then leaves the room. While she is gone an object is moved. The child is then asked "where in the picture is the object?"

Most three year olds fail this task. This suggests, initially, that they lack a competency with the concept of representation. After all the photograph, the map and Sally's belief are all representations which go out of date and, consequently, diverge from reality. If this is correct then three year olds are failing to understand that a representation, concrete or pictorial, can be false. If the ability to pass the false belief test is based in an ability to understand representations then the four year old should have no trouble with the false photograph and map tests, which test the ability to determine when a

---

21 Although see (Baron-Cohen 1986) for a sequencing experiment designed to distinguish behavioural from intentional interpretations.

22 (Zaitchek 1990)
representation has gone out of date. However, four year old children do better at the false belief test than the false photo and map tests, at which their performances are not significantly better than three year olds.\textsuperscript{23}

This suggests that the ability acquired at the age of around three and a half which enables a child to pass the false belief test is not the ability to understand representation in general but the ability to interpret intentional behaviour. This underlies the claim of ToMM theorists that the ToMM module is specifically linked to the ability to attribute mental states which can diverge from reality, rather than a specific application, to the behavioural realm, of a general cognitive mechanism.

There is definitely a difference in the way three and four year olds interpret behaviour. The three year old appears to explain others' behaviour in terms of states which simply reflect reality, or at least the situation as perceived by the child. This is true for first as well as third person beliefs. Gopnik reports an experiment in which children watch a dolls' house which has a doll inside it. The experimenter the opens the dolls' house, takes out the doll and replaces it with a teddy bear. Children are then asked "When you first saw the dolls' house what was inside it?" Only those who answer correctly, "a doll", are allowed to continue in the experiment. The question establishes that the children can correctly identify the past state of the house. However when the children are asked "what did you first think was in the dolls' house?" they reply "A teddy bear." This experiment has been replicated using a number of different formulations of the question and different techniques to ensure that the children are not misremembering or being confused by the way the question is put. All these experiments confirm that, while the child can correctly identify past situations and changes to that situation (the children correctly report that the doll has been replaced by the bear) they cannot correctly identify their own immediately past beliefs. When three year olds are

\textsuperscript{23} The results of these experiments are summarised in support of ToMM hypothesis by Leslie (Leslie 1991 pp95ff)
asked to identify their own beliefs they make the same mistake they make in identifying the beliefs of others; they treat them as simple mirrors of external reality. Four year olds, however, correctly report that when they first saw the doll's house they thought there was a doll in it.

This result, in combination with the false photo and false map experiment, indicates that the ability to understand changes in mental states, first and third person, is distinct from the ability to understand changes in physical reality or external representations of physical reality.

3.4. Theory of Mind: Constructed or Innate?

It is important to remember that these theoretical models are thought, by ToMM and CTM theorists to be known implicitly. By implicit knowledge Gopnik means knowledge which does not depend on conscious inference. Gopnik uses this implicit-explicit distinction, in the same way as Churchland, to support the claim that perceptual and other cognitive judgements are informed by theory. Recall the discussion at 2.2.2 of Churchland's claim that early in life we acquire the theory of intentional psychology which functions implicitly in our judgements about the reasons for other people's behaviour. We naturally and automatically assume that other people act in the way they do because of the typical connections between their internal states and their behaviour. But this assumption is actually the product of a theory. We are like ultrasound diagnosticians who have been trained to "see" abnormalities on the screen. After many years' experience they no longer have to make an explicit inference from the scene before their eyes to the probable cause. Rather, they report their beliefs in directly perceptual terms. This is the illusion of expertise. Gopnik's account is intended to provide the empirical support for Churchland's contention by showing how we acquire the illusion of intentional expertise through our mastery of an intentional theory. The claim, distinctive of CTM, is that the theoretical models are developed as a result of the incapacity of
previous theories to cope with the evidence presented to the infant in the form of other people's behaviour.

This claim is highly implausible, for reasons which will become apparent in the next chapter, which discusses autism. An a priori reason for suspecting the claim is that the theoretical models employed by children succeed each other in a predictable sequence. It is reasonable to assume that the evidence, the behaviour of the other people it encounters, does not change, yet the child develops a series of theoretical models interpreting this behaviour. It seems likely then that the succession of theoretical models is not the result of a series of encounters with disconfirming evidence. Rather it seems more likely that the child's cognitive abilities change over this period.

This intuitive result is not necessarily in conflict with CTM. However, empirical CTM will claim that any cognitive changes are general, equally applicable to different domains, and not confined to the child's ability to read minds. However, if inferring others' mental states depends on a general cognitive ability, equally applicable to mental states, the physical world and representations thereof, then the discrepancy between the False Belief and False Photo tests is an anomaly which empirical CTM needs to explain.

ToMM explains the succession of theories employed in mind reading as the progressive modification of a series of precursor states of a discrete module whose purpose is to infer the mental states of others. The problem posed by the succession of theoretical models is thus slightly different for ToMM. ToMM needs to show that one theoretical model arises from modifications to its precursor state in the ToMM module. Providing this can be done, ToMM has at least two explanatory advantages over CTM. The first is that it is possible that a state which is not theoretical at all may nevertheless be a precursor of a state which is. This is important because newborns have fairly unsophisticated cognitive abilities.

24 I am not denying that the child is treated in different and more complex ways as it grows. To that extent the evidential base changes. I am merely pointing out that, between, say, the ages of three and four the child does not seem to acquire any new evidence which might falsify her three year old conception of other minds. People do not suddenly start behaving 'opaquely' rather than 'transparently' in their interactions with children once they turn three.
abilities. We may not want to say that the baby who attracts its mother’s attention by crying has a theory that the mother loves her and wants to comfort her. Nonetheless that ability, and the reciprocal engagement with the mother, may perhaps be precursors of more sophisticated abilities which develop later and which are more plausibly explained as instances of theoretical mastery. Empirical CTM however has to explain the relation between newborn and later abilities in terms of the succession of theoretical hypotheses.

The second advantage ToMM has is domain specificity. As I said earlier, ToMM relates only to the ability to read minds. Thus it is not an embarrassment for ToMM that three year olds who already make correct inferences about physical reality and representations do not infer the presence of mental states. ToMM can treat cognition involving physical reality and representation as discrete cognitive abilities handled by another module, or as precursive in the development of the full theory evidenced by passing the false belief test.

One clear difference between empirical CTM and ToMM is in the explanation of developmental deviance. Consider the case where a child is unable to pass the false belief test even at the age of five or six. If such a child had developed all her other cognitive abilities at the normal rate the natural conclusion would to be that her impairment was specific to the recognition of belief. The ToMM hypothesis would be that the ToMM module or a precursor stage of it had been damaged. The CTM hypothesis would have to be that the child was perhaps evidentially impoverished with regard to belief, perhaps through lack of social interaction.

Such a case would favour the ToMM interpretation, particularly if it could be shown that the child also lacked a precursor of full ToMM. However there are other cases which might favour the empirical CTM interpretation. Consider a child with Down’s syndrome. If such a child also failed the false belief test this would tend to support the view that the acquisition of intentional
Chapter 3  The Child’s Theory of Mind 129

understanding was only one aspect of a general cognitive ability impaired by Down’s syndrome.\(^{25}\)

These hypotheses suggest a field of CTM research which has expanded rapidly in the last decade. However, interpretation of the evidence provided by comparison of the developmental trajectories of normal children with those with developmental delay is contested. I shall argue in the next chapter that the evidence from Autism does not unequivocally support ToMM or empirical CTM. In the rest of this chapter I show that the evidence gained from experiments with normal children is not conclusive either.

3.4.1. Early Stages in the Development of the Child’s Theory of Mind?

Belief is the last basic intentional concept acquired by the child at about the age of three and a half. Only with this concept in place is the child able to reason using the practical syllogism. The components of the practical syllogism refer to aspects of other persons’ “internal subjective lives”\(^{26}\) and their connections with action. However the Theory of Mind hypothesis derives part of its attractiveness, not just from its ability to explain the activities of four year olds but because of its explanation of stages of development prior to the age of four.

The classic theory of infant cognition, proposed by Piaget, did not regard the very young child as able to make the self/other distinction in any form. According to Piaget the infant is a solipsist whose experience of its own sensations, perceptions of objects in the environment, and of other people, are all understood as events in the child’s mind, which is not conceived of as distinct from the rest of the world: effectively, the child’s mind is the world. The awareness that there are other people with subjective experiences distinct from

\(^{25}\)Assuming that the only impairment in Down’s syndrome is general cognition. Given that Down’s is also correlated with a range of developmental disorders it might not be surprising if the ToMM module was affected. In such a case if mindreading was delayed or absent it might look as if it was the result of the general cognitive delay resulting from Down’s syndrome when in fact it was the result of a more specific impairment to the ToMM module.

\(^{26}\)(Wellman 1993)
those of the child is an epistemological achievement, not completed until the age of seven.

However, there is now evidence that the child is born, if not with concepts of the self, of others and their mental states, at least with some fairly rich precursors. Infants pay attention to other people and their movements in ways which suggest that they are innately predisposed to perceive other humans differently from the rest of their environment, not only as different from the rest of the objects in the environment, but as beings of a similar sort to the infant. There is evidence that children are innately disposed to recognise faces, and perhaps expressions like smiles, so that their attention automatically turns to other people, particularly the mother in the first few months after birth.

However, face recognition only explains the preferential attention of infants to other humans, not an awareness of the self-other distinction or an understanding that others’ expressions are produced by their subjective experiences. Further evidence for an innate awareness of the self-other distinction comes from the infant’s ability to imitate the facial movements of others. Imitation requires recognition of others’ expressions and patterns of movement and control of one’s own body. One proposal is that infants are innately disposed to interact with others by imitating and that such imitating indicates the ability to recognise that other bodies can be controlled. It also indicates an awareness of subjective experience and its connection with bodily control — awareness that the bodies who are engaging in games of imitation are doing so on the basis of experience, which is distinct from that of the infant. The idea is that games of imitation, while they may not require fully fledged concepts of self and other, do require the awareness that the other has subjective states which play a role in relation to bodily movement.

Infants are never strict behaviourists: one fundamental assumption of mentalism — that external visible behaviours are mapped on to phenomenologically mental states — is apparently given innately. Clearly infants have much to learn about the
Another striking fact about games of imitation is that they are a source of pleasure for parent and child. Both parents and newborn children actively enjoy these games, smiling and making noises when imitation is successful. Perhaps these games communicate enjoyment; that is, the infant smiles in order to make its mother smile and vice versa. If that is the case then there is quite a sophisticated Gricean ability in play; if not the ability to recognise something as specific as intentions and meanings of gestures at least the ability to understand that one’s movements can succeed in bringing about subjective, pleasurable states of the other.

If imitation is innate (and the earliest instance observed by Meltzoff and Moore was forty two minutes after birth) and involves awareness of others’ subjective experience, that awareness is based on representations of the infant’s body state. On the first person side of the self - other divide the relevant subjective experience is proprioceptive body control. On the other side of the divide, the other’s experiences are not conceived of as states directed to specific objects: states such as desire for a drink, seeing or believing that the other person is happy, or intending to produce a smile or protruding tongue. Rather the infant is merely aware that the other controls its body. At the earliest stage it would be merely an awareness that the other is in a state 'like mine' whose production is the goal of the imitation game.

Such abilities link proprioception and perception to action via the representations involved in autonomic body control. They do not require the

---

27(Gopnik 1993p340)
28This, of course, is a general point about the representations involved in proprioception and basic body control. One might also say the same thing about the representations involved in the monitoring of basic visceral sensation and body function. I discuss these representations in the section on Dasmacio in the next chapter. At this point we can note that the point about the representations involved in proprioception is similar to that made by Evans in his discussion of self reference. He argued that there was a way of referring to oneself in thought in which the question of the identity of the referent did not arise: namely those thoughts involved in bodily orientation and awareness. He called this aspect of first person thought Immunity to Misidentification. In the present case the difference between M-representation and Evans first
explicit taking of attitudes to propositions which express concepts like "self" and "other." Nonetheless, imitation games do involve intensive engagement with another person. If the basis of such engagement could be extended to include, not merely bodily movement, but recognition of external objects or other people that would indicate a further development in infant cognition: awareness that others' subjective experience can be directed outwards rather than being confined to the monitoring and control of the self.

One source of the chimpanzee theory of mind was the sophisticated use chimpanzees and other primates seem to make of a concept of others' visual attention, both in soliciting alliances and in tactical deception. The chimpanzee theory of mind treated the chimpanzees' understanding of perception as a concept of an intentional state of perception or visual attention, linked with a concept of intention to explain behaviour. Young infants are not this sophisticated but by the age of nine months they do engage with other people by monitoring their people's gaze (as do the chimpanzees in acts of concealment and deception). Infants will look intently into others' eyes and if the other looks somewhere else they too will look in that direction. They are able to play this game in reverse, looking first into the other's eyes, then at a salient object then back to the other's face to check that they have been successful in diverting the other's gaze.

Around this time children also start to comprehend pointing. They will turn and look in the direction, and to a specific object, pointed by an adult. (Interestingly, pointing seems to be a specifically human form of ostension, other species, like piping plovers, monitor gaze, and chimpanzees will take an experimenter by the hand and lead her to a desired object if she does not follow the chimpanzee's gaze, but no other animals point to salient objects.)

person thoughts is that, for the former the possibility of misidentification does arise, precisely because they express an attitude to a proposition, whereas the first person thoughts with which Evans is concerned do not require the taking of attitudes. See (Evans 1984) chapter seven
From this sort of evidence, Baron-Cohen has argued that children have the concept of attention. That is, they are aware that their own and others’ subjective experience can be directed to objects in their environment. There is some evidence that as early as two months children are disposed to "point" at things they are looking at. This of course proves little except that the possibly innate human ability to point is being rehearsed even at this early stage. Such gestures do not indicate an awareness that the others' subjective experience is being engaged by the act of pointing.

For the same reason so called proto-imperative pointing in which the child points at things it wants is not conclusive. It needs to be shown that the child knows that in pointing at something she will divert the attention of another person to it rather than merely cause the other person to give the object to the baby. This is another instance where a sequence of actions — pointing, looking, fetching — can be interpreted as a purely physical sequence or in intentional terms such as attracting the other's attention, and informing her of the agent's desire for the object.

Proto declarative pointing, on the other hand, is difficult to interpret without invoking the awareness of others' subjective experience of the objects pointed to. Here the baby points to objects, not to indicate that she wants them but to make them an object of joint attention or response. The baby points to a toy or the source of a noise and then looks at the face of the parent to see how she responds. The pointing is intended not just to make the other aware of the object but to "comment on it as a topic of interest, concern, fun etc." 29

Further experiments with younger children indicate that when they are unsure of a situation (the meaning of a gesture or the role of an object they have been given) they look first to the adult's eyes.

From these results Baron-Cohen concludes that the ability to monitor another's gaze, to rely on it for information and as a source of reassurance of joint involvement in activity indicates an understanding of attention and the

29 (Baron-Cohen 1993p73)
ability of subjects to make an object the focus of joint attention. He proposes this as a precursor stage in the development of full ToMM. This hypothesis is derived from small samples but it is consistent with other results, particularly those with autistic children (which are discussed in the section devoted to that topic in the next chapter and briefly in the section on precursor states below).

The evidence does seem to show that children (like autistic children) who subsequently fail to develop full mindreading skills are those who, even at this early stage, do not involve themselves with other people by engaging their attention. That attentional engagement should be so important in the early development of the understanding of mind is not surprising. It fits very well with findings about the role of gaze and attention monitoring in primate psychology.

The ability to share and direct attention by the use of gaze and pointing suggests that at the age of one the child understands that experience is subjective in the sense that distinct subjects confront an independent world on the basis of what they see. So experience is intentional in at least the sense that it can be directed toward a mind-independent world. However I have been using intentional in a more specific sense as a thesis about the content of experience and its objects. Experience is intentional in the sense described at 1.2 if it is "about" its object in virtue of its content, in the same way as a picture or sentence is about its object.

3.4.2. The Understanding of Opacity in Early Childhood

ToMM theorists accept this philosophers' account of intentionality. They are interested in the developing ability of the child, not just to direct its attention, but to distinguish between transparent and opaque mental states. The two precursor states I have considered so far are both transparent. When the child makes the self-other distinction we can say of her "The baby sees its mother smile". Any other description of the mother can be substituted without
affecting the truth of the sentence. Similarly at the stage of shared attention we can say "the baby is attending to the teddy bear". We could substitute any other description of the teddy bear and still correctly describe the baby's mental state "the baby is attending to the stuffed bear replica" would do just as well. However, the same is not true when we move from a state of attention to a state of belief. There is a difference, on an opaque reading, between "the baby believes the stuffed replica bear is in the cot" and "the baby believes the teddy bear is in the cot" even though both sentences refer to the same thing. The difference is that the concepts "stuffed replica" and "teddy bear" are distinct. The ToMM theorist explains that sort of conceptual mastery as the acquisition of concepts, understood as representations of attitudes to propositions. Thus the child can only be said to have an understanding of opaque mental state concepts when her actions indicate that she understands other people, and herself, as engaging with the physical and social environment by representing it and taking attitudes (of belief, expectation, desire) to those representations: hence the emphasis on M-representation.30

So far I have described some of the abilities a child exhibits in its social interactions up to the age of four. Self-Other, Attention, Proto-Imperatives and Declaratives, Transparent Representation and Opaque Representation. The last is mapped by the Theory of Mind theorist onto the differential abilities of three and four year olds marked by the false belief test. The way I have presented the case it looks as if these are stages a child passes through on the way to acquiring an intentional concept of belief, and this is unsurprising because this is the developmental trajectory on which ToMM theorists, interested in accounting for the false belief results, have concentrated. On this account the earlier states are precursors in the development of the concept of belief.

30 The actual mechanism of representation is not specified in CTM. This is not a defect specific to CTM, all representational theories of mind ultimately depend on the claim that the brain can acquire and process information about the world in a way which sustains the attribution of representational capacities. Exactly what type of world-brain relationship qualifies as suitably representational to qualify as an intentional mental state is the question. For a discussion of the relation between notions of representation of information and intentionality see (Sterelny 1995).
However this cannot be the complete story because a theory of mind is not just a way of conceiving of belief, but of the way a network of mental states interact to explain behaviour. So far I have said nothing about the way a child conceives of other mental states. Does the child's understanding of desire undergo a similar series of transitions? There is some evidence that it does. An analogy of the false belief test designed to determine whether children can recognise their own and others' desires seems to show that there is a similar transition from a transparent to an opaque conception of desire. Astington and Gopnik have documented a change from a transparent to an opaque concept of desire about the age of two.31 Similar transitions, occurring at different intervals after birth, have been noted for other mental states. What follows is an account of the development of an opaque understanding of psychological concepts consistent with these findings. It is intended to be an account of the developmental evidence which all accounts would agree on as explananda for which their theories are competing explanantia.

The newborn experiences drives for affective contact and nourishment. Such states are intentional in that they have conditions of satisfaction, but are not governed by representations with propositional content. Later, perhaps at the proto-declarative stage, the child understands others may have desires for amusement, affective contact, food or objects in the environment. That is why she spends so much time playing games of pointing and hide and seek with the mother. At this stage the child has a concept of desirability, of objects which are attractive, to look at or to eat. Such a concept need not involve a notion of the environment as represented in the mind of another, the child may just make the connection between sensations of satisfaction and the action of looking, grasping or eating. A more sophisticated step is the idea that objects may be thought of, or transparently represented, as desirable. Three year olds understand such a concept quite well in their own case. They understand that the desire component of a reason for action can change over a short time, either

31(Astington J. & Gopnik 1991p44ff)
because the original desire is satisfied or something more desirable takes the place of the original object. (This is an analogy of the representational change test conducted for belief in which the three year old children misidentify their own immediately past beliefs, described at 3.3.3. 32 ) Interestingly, unlike the belief case, the results are not so clear cut when the child is asked to attribute desires to others. 33 Astington and Gopnik suggest that this may be due to the inability of the three year old child to recognise others' beliefs. Finally, at about the age of three, the child can pass an equivalent of a false belief test for desire, indicating that she understands that others represent objects as desirable and act on the basis of those representations.

So the development of the fully fledged intentional understanding of desire, precedes the intentional understanding of belief. Both processes seem to undergo a similar transition, from a stage at which the child understands the subjective experiences of self and other, to a stage at which that experience is understood as directed to a world of objects, via a stage of transparent representation, to opaque representation in which objects are understood as desirable from another's perspective. 34

These results would indicate that intentional concepts are acquired in similar ways but at different rates. Most work has been conducted on the last stage in the process; the transition from a transparent to an opaque conception of the state in question. Tests on normal and Down's syndrome children indicate that the normal child arrives at an opaque conception of mental states in the following order: Perception, Imagination, Pretence, Desire, Belief. 35

---

32 (Astington J. & Gopnik 1991)
33 Interpreting results in this area, particularly the recognition of the representational nature of other's desires, is complicated. A particular problem with interpreting "false" desires is that to determine what another person wants, one needs first of all to know what they believe. This is a product of the holistic nature of mental state attribution. So if a child makes a faulty prediction about what another person will do this may be due to her misattribution of a desire or a belief. The problems are discussed by Gopnik and Astington.
34 Wellman puts the change from a transparent to an opaque conception of desire at about the age of two and a half (Wellman 1993)
35 This result relies on the following methodological assumption: relative degree of difficulty in passing analogues of the false belief test equates to order of acquisition of the concept in question. So if a smaller percentage of three year olds pass the false belief test than pass a false desire test, experimenters assume that children understand the concept of desire before they understand the concept of belief.
As I described it above, the understanding of the concept of belief goes through precursor stages of Self-Other, Attention, Proto-Imperatives and Declaratives, Transparency and, finally, Opacity. Does this sequence occur independently of similar sequences for other intentional states? There are at least two versions of the hypothesis that the transition to an opaque concept of belief is not independent of stages in the mastery of other intentional concepts.

Wellman has proposed that the child is a desire psychologist before she is a belief psychologist. The opaque understanding of desire is a precursor in the opaque understanding of belief. The idea is plausible because the child is constantly confronted with instances of unsatisfied desire, for affection, for attention, for nourishment and objects to play with. So the idea that one can be in a state of mind whose conditions of satisfaction are not met is one which the child encounters early as it tries to satisfy its desires. This understanding of desire is progressively developed through a series of precursor stages which culminate in the opaque understanding of desire at around the age of three. The child passes from explaining others' actions in terms of objects desired by the child, to objects desired by the other, to representations of those objects entertained in the mind of the other. With this opaque understanding of desire in place the child can extend that conception to the other key concept in intentional understanding: belief.

This is particularly plausible on a model congenial to CTM. The child constructs a simple hypothesis, progressively makes it more sophisticated and finally extends it to cover mental states other than the one for which it was initially developed. However, the idea can equally be incorporated into ToMM. This involves the abandonment of the idea that intentional understanding of each concept develops independently through a series of precursor stages at different rates, and incorporating the idea that the last stage in the opaque understanding of desire is the penultimate precursor in the understanding of belief.
3.4.3. Understanding the Concept of Pretence.

Another hypothesis is that the precursor stage in the transition to an opaque conception of belief is pretence. Children from the age of two are able to engage in games of pretence and deception. Pretence comes in many forms: object substitution, deceptive behaviour, games of cowboys and Indians. The first, object substitution, does not imply a knowledge of opaque mental states in others. If a two year old child talks into a banana pretending it is a telephone then the child imaginatively suspends the evidence of perception for the sake of the game. This act of imagination is confined to the first person. However, for the child to understand another's act of pretence, the child must be able to understand that the other person is also imaginatively suspending the normal understanding of the environment, based on the evidence of veridical perception. On Leslie's account, to deceive another person by pretending, the child must be able to understand that another would form false beliefs on the basis of the information conveyed by the act of pretence. Consider the simplest example, a child pretends to drink from an empty cup and then gives the cup to its mother who also pretends to drink. For the cup to serve as a prop in this game of make believe the child must know that the mother is also pretending that the cup is full and is pretending to drink. These acts of pretence cannot take place on the basis of what is perceived, so both participants in the game, are, for the purpose of the game, attributing mental states other than the transparent beliefs which normally explain action.

Analysing the nature of the mental states involved in games of make believe is difficult. Are the baby and the mother acting on the basis of a false belief (that the cup is full), or is pretence not a matter of entertaining false beliefs but of suspending the attitude of belief? These questions are often discussed in analysing our understanding of fictional works. For example, is it true simpliciter that "Sherlock Holmes lived in Baker Street", or is it true relative to the fiction of Conan Doyle, or is that sentence neither true nor false
but merely consistent with some of the other statements made in Conan Doyle’s work? How one answers this question will have some bearing on what one says about the attitudes of someone who reads Conan Doyle. Are they forming false beliefs? Beliefs true to the fiction? Are they pretending for the purposes of the fiction? It is not my purpose to answer this question but to note that when people engage in activities of pretence or make believe these questions arise.

Leslie proposed that pretence takes its place as a precursor stage in the development of a Gricean theory of communication. The Gricean theory requires that we understand another’s meaning by inferring her intentions, part of which requires inferring her beliefs. The use of proto-declaratives requires joint awareness of the others’ mental states, not only of attention, but surprise or pleasure. A pointing gesture is communicative because the recipient must infer the other’s mental state in order to understand the import of the gesture. However, proto-declarative communication can be understood without the attribution of an opaque understanding of mental states to the maker of the gesture. One can infer that the maker of the gesture is surprised or pleased without understanding that the maker of the gesture is opaqueely representing the object as pleasurable or surprising.

Leslie proposed that simple games of object substitution (riding the broomstick as if it were a horse) and pretence marked a cognitive advance towards the ability to recognise others’ opaque mental states. According to Leslie when the child understands that the mother is pretending to drink from a cup the child has to "decouple" her transparent perceptual understanding of the situation in order to allow it to be overridden, for the sake of the game, by the pretend scenario. Pretence requires both pretenders to engage in this act of decoupling.

If this is correct, then pretend play does require an understanding of an opaque concept of pretence. There is however, a stronger, version of this argument which adds a hypothesis about the nature of opaque understanding itself: that it consists in the ability to metarepresent, or form second-order
mental representations of others' first-order opaque mental states. There are good reasons for resisting this conclusion. The ability to metarepresent, or form second-order representations of first order representational relations is a generalised cognitive ability. For any first order representation the child who can metarepresent should be able to understand the nature of that first order representational relation. So the child who can metarepresent should be able to understand the relation between desires and their objects, perceptions and their objects, and pictures or even sentences and their objects. However the only representational relation the child can understand at this stage is pretence. Yet the idea, for example, that another may find an object, which the child does not desire, desirable, does not seem any more difficult than understanding the idea that the other is pretending an empty cup is full. Both acts of understanding require the child to understand that the other is representing an object in a certain way. An opaque understanding of belief may be slightly harder, but a child who can metarepresent should have no difficulty with the concept of false belief especially if we analyse the concept of pretence as the entertaining of a false belief within a limited context. Yet the fact remains that the normal child who understands pretence at the age of two does not understand desireopaquely until the age of three and does not understand belief until between three and four. Pretence is an island of precocity in the child's intentional repertoire.

There are three conclusions I would like to draw from the discussion of pretence: The first is that pretence is not to be analysed as a false belief relativised to a particular situation, but as a different and sui generis propositional attitude.

The second is that understanding pretence does not require a general ability to metarepresent, only the ability to understand the concept of pretence.

The third and most important is Leslie's insight about the communicative role of intentional understanding. The child's developing theory of mind is employed within a very intense communicative environment and is primarily a
matter of reading meaning into the gestures, expressions and vocalisation of the other. So Leslie is correct to connect pretend play with protodeclarative gesturing. If this is correct, and the abilities which ToMM seeks to explain are primarily social rather than an aspect of cognition in general, it is not surprising that the child does better at predicting and interpreting other people than at handling abstract operations such as manipulating photographic representations.

3.5. A Note on the Concept of Precursor States

ToMM proposes that the ability to form M-representations is the last stage in the development of a module devoted to interpersonal understanding. Clearly the infant is not born with the ability to form M-representations, so the question of how it develops that ability naturally arises. One answer to this question is suggested by evidence that children reciprocally engage with others from birth and that there is a predictable sequence in which babies and children exhibit distinct types of intersubjective behaviour. Examples I mentioned earlier are: preferential attention to faces, imitation, shared attention and gaze monitoring, games of pretence, the transparent understanding of perception, desire and knowledge, and, possibly, the opaque understanding of such concepts, underwritten by the ability to form M-representations.

These abilities are typically manifested in a sequence which invites the hypothesis that some of the earlier abilities are precursors of the later. But which? Is the hypothesis of ToMM that the mechanism involved in early imitation or face recognition is steadily transformed into the mature device which forms M-representations in order to explain others’ behaviour? Sometimes, as in Wellman’s earliest formulations of the theory theory, it looks as if this is the case. However not all the abilities exhibited by infants and young children can be preclusive in this way because some of them, such as the ability to detect and follow gaze, can be absent without affecting the later
development of mindreading skills. People blind from birth nevertheless develop excellent mindreading skills. So the issue of precursor states raises two questions: which states are precursive of the others, and how does the transformation occur? Most of the discussion of these issues focuses on the final transition, the one marked out by ability to pass the false belief test.

Precursor is, unfortunately, a vague term. It suggests that the later abilities depend on the instantiation of the former, without specifying the nature of that dependency. In the loosest sense of the term one event or process is a precursor of another if it is temporally prior and non accidentally connected to a later one. In this loose sense John the Baptist is a precursor of Jesus because he came before Jesus and announced "There cometh one mightier than I after me". However the mere fact that events typically occur in a predictable temporal order or sequence is something which invites explanation, rather than explaining the relation between those events.

So, to the discerning child psychologist, the manifestation in a child of early imitation or joint attention announces that the child will, in all likelihood, go on to develop the ability to attribute intentional states. So these states are precursive of intentional understanding in the loose sense just discussed. However, the discerning psychologist will not be content merely to observe the correlation and make the appropriate predictions. Consider again the analogy with language acquisition. A child who starts to say "I runned" will normally develop the ability to use the past tense correctly. So the use of expressions like "I runned" is precursive of the ability to conjugate regular and irregular verbs. A likely explanation is that the production of expressions like "I runned" is the result of the application of the linguistic rule that verbs must agree in tense and number with the noun-phrase with which they are correlated. The ability to parse and produce sentences in this way comes "on line" in children at about the age of three, and is first seen in the use of expressions like "runned" and later fine-tuned to extend to irregular cases. In this case the first stage is a precursor of the second because they both depend causally on successive stages
in the activation of the same module. With this notion of precursor in place, psycholinguists hypothesise that the understanding of the past tense may be the result of changes in the circuitry of certain areas of the brain (Broca's or Wernicke's regions), changes whose occurrence is determined by that part of the genome which controls the development of these areas.

An analogous case in the case of intentional understanding might be that the same brain circuitry is genetically programmed, firstly to construct representations which enable joint attention, and, subsequently, to construct M-representations. If this is correct then joint attention is a precursor of full intentional understanding because the representations on which both abilities depend are sustained by the same neural circuitry and the transition between joint attention and full intentional understanding depends on genetically driven changes in that circuitry which sustain specific types of representation. For reasons which will become apparent in the next chapter, I think that this account of the relation between joint attention and intentional understanding is incorrect, but the notion of a precursor it exploits is a coherent one and susceptible of empirical verification.

Note that this notion of a precursor treats joint attention and intentional understanding as underwritten by the same mental mechanism. This type of precursive relationship has been described as "the tadpole and the frog" model of precursor states, presumably because the later stage is a development of its predecessor, just as the frog develops from the tadpole. This is a slightly misleading way to describe the situation, because any earlier stage of an organism is a precursor of a later stage. Since we don't want the relation between the states to be a trivial one we should restrict it to genetically driven changes in modular capacities. (There are, of course, changes to a modular capacity which are not genetically driven. Things like kidney disease or hepatitis disrupt modular organic functions. The early stages of such diseases are precursive of the later ones)

36Firstly by Gomez in his admonitory piece on the overly-optimistic attribution of ToMM capacities to primates. See(Gomez 1991)
This way of characterising the notion of a precursor allows a significant, nontrivial, contrast with the "lock and key" hypothesis, which allows that changes in a precursive mechanism may cause changes in a distinct mechanism. An example might be the overproduction of a certain hormone or neurotransmitter as the result of malfunction in one area of the brain, which then disrupts the functioning of another module. If joint attention and intentional understanding are indeed separate modular capacities then, if changes in the joint attention module typically bring about changes in the intentional understanding module, their precursive relation will be an example of the lock and key hypothesis.

The usual locks and keys are chemical/structural relationships. As the main proponent of the hypothesis says: "Lock and Key hypotheses are of course widespread in biology: for example DNA base pairs can only lock together in a specific way; antibodies can only fit on the surface of specific antigens...", 37 However, while the lock and key hypothesis is plausible for such chemical interactions, since it describes a physical structure, the relation between the abilities exhibited by young children is more difficult to accommodate to this hypothesis.

There are, however, at least two interpretations of the development of intentional concepts in children which employ a version of the "Lock and Key" hypothesis. The first is Leslie's, discussed earlier, which treats pretence as a vital precursor in the development of an opaque conception of belief. The second, closely related, is Simon Baron-Cohen's proposal that the representations which underlie shared attention "unlock" the M-representations on which the four year old relies in passing the false belief test.

For Leslie pretence unlocks belief because pretence requires the pretender to take an attitude to a proposition which is false in the real world but true for the purposes of the game. Pretending requires P-representations which represent an object, a proposition, which is false of that object, an agent

37 (Baron Cohen & Swettenham 1996)
and the agent's attitude to the proposition. Belief requires an M-representation with an attitude of belief in the attitude slot. The step from understanding pretence to understanding belief is the realisation by the child that all actions, not merely those involving pretence can be explained in terms of attitudes to propositions whose content is not necessarily veridical. This type of lock and key hypothesis treats the opacity of the representation involved in pretence as a key, because both P and M representation have slots for agents, attitudes and propositions whose content can diverge from reality as perceived.

Baron-Cohen has made a similar claim for the relation between joint attention and belief. He believes that sharing of attention is the precursor to M-representation because shared attention requires the agent to represent another person as "perceiving" that a certain proposition obtains. Shared Attention Mechanism (SAM) representations have slots for agents, propositions, and fill the attitude slot, occupied by Pretence in P-representation, with a transparent attitude of perception. Transparent, because in sharing attention the child treats the other as perceiving the world in the same way as herself.

There is a contrast between Baron-Cohen and Leslie over the nature of the "key" involved. According to Leslie it is the opacity of the precursor attitude of pretence, according to Baron-Cohen it is the triadic structure of SAM. There is some empirical support for both theories in the study of autism, discussed more fully in the next chapter. Autistic children and adults are paradigmatically unable to "read minds". Interestingly one diagnostic test for autism is the failure to engage in games of pretence. Of course this is not conclusive, the failure to develop mindreading ability and the failure to pretend may have a common underlying cause (perhaps in a failure of an appropriately directed imaginative capacity as Simulation Theorists suggest38) rather than stand in a precursive relationship. Nonetheless failure to understand the concepts of pretence and deception is a very reliable indicator of autism.

38 See especially (Currie forthcoming)
However failure of SAM is emerging as an equally good, if not better, predictor of the failure of autistics to develop full mindreading. Baron-Cohen reports the results of tests on 16000 children at the ages of 18 and 19 months for the presence of pretend play, protodeclarative pointing and shared attention. Only 14 children failed all three tests. Of these 14, 10 were later diagnosed as autistic. All these autistic children failed two SAM tests. Baron-Cohen reasonably concludes "a severe impairment in SAM is an important indicator in the early diagnosis of autism".

It is interesting that failure of pretend play and SAM appear to go together. Baron-Cohen reports no instances of total failure of joint attention accompanied by the presence of pretence and none of total failure of pretence with intact SAM.

These facts suggest that pretence and SAM are both connected with mindreading for one of two reasons. The first is that mindreading is the development of a mechanism which has two characteristic and roughly contemporaneous manifestations in SAM and Pretend play. Thus they both develop from a joint precursor or rely on a mechanism which is precursive of mindreading. Since SAM explains the attributions of transparent intentional states and Pretence is allegedly understood by the child as an opaque intentional state, opacity cannot be the key. Baron-Cohen suggests that triadic structure may be the "key" which unlocks M-representation. However the correlations observed by Baron-Cohen do not suggest that triadic structure of the propositions represented by children, any more than the capacity to represent concepts opaquely, is essential to understanding other minds. These properties are artifacts of the philosophical discussion of the logical properties of the sentences which talk about psychological states. (Recall the discussion of opacity at 1.3.) There is no necessary connection between the structure of propositions which ascribe intentional states and the way the actual representations in the mind of the agent are structured. This merely reiterates

39 (Baron- Cohen & Swettenham 1996)
the point made in the discussion of the Game of Life at 1.5 and again at 1.8. The
same point arose in the interpretation of the chimpanzees "deceptive"
behaviour. Recall the chimpanzee who hid and watched until the his
companion, who had avoided eating until the first chimpanzee had left the
scene, revealed the food, and then returned to steal and eat it. There is, no
doubt, a description of this behaviour in nonintentional terms but it seems
quite reasonable to say that the first chimanzee "pretended not to know "
where the food was and that the second "pretended to be deceived". Ryleans
do not object to describing the chimpanzees' activity in this way because for
them intentional language interprets behaviour as meaningful without
expressing any hypothesis about the nature of the processes which underlie
that behaviour. One can use these terms without any commitment to the view
that the representations involved have anything like the structure of the
multiple embeddings needed to render the logical form of "pretended to
believe the pretence". Chimpanzees understand many things about each other's
behaviour but that understanding does not depend on the operation of a parser
for embedded propositional attitude sentences.40(See also note 4.)

Similarly with children. They understand many things about human
behaviour, but that understanding does not necessitate the presence of
representations whose structure mirrors that of the sentences we use to explain
the child's behaviour. So if pretence and SAM are manifestations of a
precursive state which unlocks mindreading, the key is unlikely to be either
opacity or triadic structure of mental representations.

One plausible account which could explain the occurrence of pretence and
SAM is that both depend on imagination. The hypothesis is suggested by
Wing's presentation of the autistic triad and has been developed as an
alternative to the theory theory by Simulation theorists such as Goldman,

40For a detailed and convincing deflationary interpretation of a range of experimental
and observational evidence adduced in favour of a chimpanzee theory of mind see especially
(Heyes 1993) Dennett is another who has consistently urged a deflationary interpretation of the
evidence for a chimpanzee theory of mind.
Gordon and Currie.\textsuperscript{41} The basic idea is that understanding other people is not a matter of making inferences from observable behaviour to theoretically postulated inner states which explain that behaviour. Rather, when we wish to know what other people would do or think in a specific situation we simply imagine ourselves in that same situation. If we want to know what someone else would do when standing on the edge of a cliff, we imagine ourselves looking over the edge of that cliff. If part of that imaginative scenario includes reeling back with fear or creeping forward out of curiosity, we attribute the appropriate state of fear or curiosity to the other person.

Obviously the representations involved in pretend play are very close to those involved in this sort of imaginative enactment since they both depend on imagining that one is in a different situation. It may well be true that the sharing of empathy and direction of attention involved in SAM depend on this sort of imaginative identification. Simulation theory has its own problems\textsuperscript{42} as a general account of the last stage of mindreading but the idea that SAM and pretence are both underwritten by some sort of imaginative ability is not straightforwardly inconsistent.

This is not to endorse Simulation theory, merely to point out that the notion of imagination or empathetic reenactment in specific situations such as joint attention or play could serve as a "key" according to the lock and key hypothesis. The trouble is that the notion of a "lock and key" precursor state is insufficiently developed by those who employ it. One can see this if one reflects that imagination might equally well be invoked by those who favour the "tadpole and the frog" hypothesis. They might claim that the child's imaginative capacities are transformed into the ability to attribute mental states.\textsuperscript{43}

\begin{footnotesize}
\textsuperscript{41}A survey of the debate between theory theorists and simulation theorists is contained in two symposia in \textit{Behavioural and Brain Sciences} (1993 Vol.16) and \textit{Mind and Language1992.Vol.7) See also (Currie forthcoming)
\textsuperscript{42}(Nichols forthcoming)
\textsuperscript{43}A version of this way of understanding imagination is present though not articulated as such in (Cohen 1993)
\end{footnotesize}
The conception of intentional interpretation I prefer does not attempt to integrate the abilities of the child by treating them as joined together in a precursive hierarchy leading to M-representation or Simulation, nor to link them as maturational stages in the development of the same module. In the next chapter I shall develop this alternative model while discussing the interpretations of deviant and delayed development in interpersonal understanding. This alternative model treats interpersonal understanding as depending on many different abilities, including those evidenced in early childhood, such as face recognition, emotional response, understanding pretence and perception. It does not, however, seek to integrate these various abilities: it just assumes that some of them are in play during any act of interpersonal understanding. If there is any "key" to intentional understanding it is, as Leslie says, in the fact that it develops in communicative situations. We should therefore pay attention to those representations needed to understand another's communicative acts. In the next chapter I offer an account of the way communication is understood which accounts for the communicative and intentional deficits of autism previously explained in terms of the failure to develop a theory of mind.

3.6. Conclusion: Precursor to later Chapters

The discussion in this chapter shows that, despite appearances, the essential issue between theorists like Gopnik and Baron-Cohen is not the theoreticity, innateness, or modularity of the capacity for understanding other minds but its domain specificity. Gopnik explains mindreading as the focusing, on a set of problems posed to the developing child by interpersonal contexts, of a general capacity for forming causal hypotheses. That is why the child is described as a "little scientist". It would not be inconsistent with Gopnik's story if this capacity for theory formation was innate and modular. Baron-Cohen, however, treats mindreading as a discrete form of cognition involving a set of intentional
concepts which are developed and applied solely in the interpretation of behaviour. They are not, for example, a subset of a general set of explanatory concepts like "cause" "effect" and "law." Like Baron-Cohen I agree that the ability to read minds is domain specific. In general domain specificity and modularity go together but there is no necessary connection between the two. Thus, establishing domain specificity does not resolve the issues of theoreticity and modularity of the representational device(s?) on which the tacit knowledge required for mindreading depends. In the next two chapters I develop an account of the tacit knowledge required for mindreading which treats it as neither theoretical nor modularised, but the arguments against theoreticity in chapter four need to be distinguished from the (I hope) persuasive considerations against modularity given in chapter five.

The reasons are, firstly, that a nontheoretical ability could be modular, as is the case with many knowhows like those discussed in chapter two, and secondly, that a set of separate modular abilities could yet be integrated theoretically. So if I succeed in showing in chapters four and five that the information required for mindreading is not theoretically unified I will not thereby have shown that it is not modular. However, in chapter five, I develop an account which explains mindreading as the integration of a set of distinct modularised capacities rather than as the functioning of a single "other minds" module. So I provide an account of mindreading as a nontheoretical, nonmodularised, ability but the arguments against theoreticity and those against modularity are distinct. I leave open the possibility that, as an empirical matter, either of them could be wrong without affecting the other.

The issue of theoreticity is a matter of integrating an explanatory domain by the application of unifying general principles. In the mindreading case these principles are intentional concepts understood by the ToMM theorist as M-representations. My argument is that we can explain the child's ability to pass the False Belief test and to respond appropriately in communicative contexts without crediting the child with mastery of a theory of the domain of
intentional psychology which she then applies to interpret her social environment.

Nonetheless, the child has practical knowledge of the propositions of intentional psychology in that the set of abilities she employs to negotiate her social environment amount to knowledge of the theory of intentional psychology. However, if the argument in chapter two is correct, this is not the same thing as unconscious intellectual knowledge of the theory of intentional psychology.

None of this, of course, speaks to the issue of modularity, which is the issue of informational encapsulation. So if I am wrong, and the child does explicitly represent some theoretical principles in order to explain behaviour, it requires a further step to show that the child does so via a modular capacity. However, ToMM theorists such as Leslie and Baron-Cohen seem to take the issue for granted by equating domain specificity, theoreticity and modularity. There is, however, as I said earlier, no necessary connection between these three issues.

In fact most of my arguments in chapters four and six are directed against theoreticity and against an automatic correlation of theoreticity with modularity. However in chapter five and again in chapter six I develop, via a discussion of some recent work in neuroscience on the way representation of emotional states affects social cognition, an account of the representations necessary for the practice of intentional interpretation consistent with the view that intentional interpretation does not depend on the functioning of a discrete module. The appearance of theoretical integration which results, through the applicability of the principles of intentional psychology to the resultant behaviour, does not signify the presence of either theoreticity or modularity at the level of the representations which cause behaviour.

Precis

This chapter discusses the claim that the developmental deficits characteristic of autism are the result of the failure, due to damage to the ToMM module, to internally represent a theory of other minds.

4.1. Autism. Preliminary remarks. Describes the autistic disability and the origin of the suggestion that it might stem from a failure to internally represent a theory of other minds.


4.2.1. The Metarepresentation Hypothesis. One interpretation of the evidence is that the difference between autistic and normal children is in their ability to form representations of representations.

4.2.2. Metarepresentation and Opacity. The Metarepresentation hypothesis is an empirical hypothesis about the nature of the mental processes on which the understanding of logical opacity depends.

4.2.3. A Dispositional Interpretation of the Evidence for Metarepresentation. The neutrality thesis applied to second order mental states.

4.3. Theory of Mind and Behaviourism. The neutrality thesis is not equivalent to a reductive behaviourism. Part of the appeal of ToMM is its association with the mentalistic side of the debate between behaviourists and mentalists.

4.4. The Autistic Dyad. Language and Social Recognition. The beginnings of an alternative account. Autism characterised as an inability to construct psychological elements of a context for the interpretation of intentional action. A context is the background information necessary to generate the inferences necessary for successful communication. The evidence relating to the typical developmental trajectory for psychological concepts reexamined to show that, from the earliest stages, the autistic child is insensitive to psychological elements of communicative
ToMM generates specific hypotheses about the way in which failure to internally represent a psychological theory using M-representations or their precursors should manifest itself in defective intentional understanding. ToMM gains much of its credence from its application to cases of developmental delay and deviance, particularly autism. ToMM, as I briefly mentioned in the section on precursors in the previous chapter, explains autism, which is paradigmatically a case of inability to read minds, in terms of damage to the ToMM module. In this chapter I evaluate those claims in more detail and prepare the way for an alternative account of the autistic deficit. The conclusions of this chapter are, first, that the case of autism shows that intentional interpretation is a matter of context construction (a context, in this sense, is the mental representation of the information needed to interpret or produce meaningful communicative acts.) Secondly, I argue that context construction is not a matter of entertaining M-representations but of entertaining the range of representations which underlie the multitracked dispositions which Ryle identified with mental states. Crucial to this activity of context construction is the entertaining of representations of affective recognition and response. Autism shows, firstly, that the affective and the cognitive aspects of context construction cannot be separated and, secondly, that M-representation is the result of an overly intellectualised conception of intentional understanding.

4.1. Autism: Preliminary Remarks

Autism was first described in a paper by Kanner in 1943 which described the inability of certain children to form emotional attachments to their parents, to take an interest in social activity and to respond to other people. Kanner was so
struck by the autistic child's inability to make contact with others, in particular to exhibit any affective responses, that he described this lack of affective response as the cardinal abnormality of autistic children and adults. However, the affective deficit, while striking, is only one of a number of abnormalities which characterise autistic behaviour. Autistic people seem to have, not just an affective deficit, but a general inability to comprehend the social world. This inability has such a variety of behavioural manifestations that a precise definition of autism remains elusive. The issue is complicated by the fact that some seventy five per cent of autistic children are more or less severely mentally retarded. In an attempt to isolate only those behavioural aspects which are attributable to autism, rather than a generalised cognitive deficit, research in the field thus concentrates on "high functioning" autistics, those whose retardation is not so severe that it prevents learning and communication. The case history below is fairly typical of the trajectory of an autistic life:

Herbert's parents became worried at his lack of social responsiveness when he was just six months old. Herbert did not make eye contact and did not smile responsively. A diagnosis of 'possible autism' was made at a relatively early age, thirteen months, because his parents persisted in seeking help with their concerns. By this time Herbert babbled a little, showed almost no response to speech, and did not point with his hand or finger to draw his parent's attention or indicate his needs. He looked at his mother only in rapid sideways glances, did not respond to her gestures, and was unable to hug or kiss...

As he grew older, Herbert became impossible to manage. If left unattended even for brief moments, he would climb on furniture, break windows and unscrew any detachable object. By the age of three he had no expressive language, and looked through or past people. He responded to his mother, who was with him all day, with hardly more recognition than usually shown a stranger. Only when he was upset would he seem to run to her intentionally.... Herbert seemed generally bewildered by the reactions of others and was unable to use other's actions to
modify his own behaviour. When he was placed in a nursery school for normal children, Herbert was unable to understand that the other children were frightened by his uncontrollable running and screeching. In turn, he was driven to states of confused excitement when his peers were engaged in active games.

By the age of seven, Herbert was clearly a very retarded child, who drifted from place to place and activity to activity with little planning. He used others, including his siblings and parents, only to satisfy his basic needs, and would involve them by taking their hands and shoving them to what he wanted. When alone, he flicked his fingers or shook his hands before his eyes, rocked and made guttural sounds and grimaced. When he was engaged in simple repetitive tasks his stereotypy decreased, and he seemed calmer. But by the time he entered puberty his face had settled into a mask-like dullness from years without normal affective responsivity.

By late adolescence, Herbert could no longer be cared for at home, and his parents prepared for the time when he would move into residential placement. They could not imagine what he would do without them. For eighteen years they had been constantly present for him — to protect him from running into the street, to try and comfort him when he was frightened by loud noises, to be always aware of where he was and to interpret for him what he needed. On the day they took him to the residential home, they were not prepared for his simply walking away, not looking back, not noticing their absence, not reflecting the meaning they presumed and hoped they had in his inner world. Herbert did not seem anxious without his parents, nor did he ever seem to miss them.¹

At the age of thirty, Herbert's life is one of quiet and orderly routines. He has no spontaneous play and no friendships, and lives a life of apparently calm structure without important others. Herbert's life history underscores, not only his long standing inability to respond to other persons, but just as poignantly, his apparent lack of desire for interaction with others and his lack of longing in their absence.

Herbert's frantic activity, his ease of excitability in the presence of others, and his

¹(Cohen 1993) This description of Herbert's history already carries within it the seeds of Cohen, Klin and Mayes psychoanalytic interpretation of autism, in its emphasis on longing and desire for the affective contact of others. Nonetheless it presents all the elements of the disorder which theorists of autism need to explain.
bewildered responses to the affective behaviours of others suggest how confusing and disorganising the social world is for him.

There are instances of autistic people of normal intelligence whose lack of social understanding does not prevent them from living and working independently. Even in these cases, however, the characteristic autistic deficit, inability to understand what others are thinking or feeling, continues to underpin autistic social behaviour. The classic expression of the autistic deficit was given by a high functioning autistic who complained that, unlike other people, he was unable to "read minds". The notion of "mind reading" with its telepathic connotation shows how mysterious the autistic finds the abilities taken for granted in human communication and interaction. As a consequence, autistics of normal intelligence cope best in situations which do not require flexible social responses such as the ability to interpret speech in context. Typically, high functioning autistics cope with the social world in a mechanical way which gives their social behaviour a stereotypical and one-sided aspect.

4.1.1. The Autistic Triad

There are at least three pervasive features of the disorder which need to be explained by any theory of autism. These three features constitute a triad because they are jointly sufficient for a diagnosis of autism. The first is a lack of social recognition, an inability to understand what is meant by others' actions and speech. This component of the triad can be seen as a precisification and extension of Kanner's original focus on the affective deficit in autism and his emphasis on autistic "aloneness". A deficit in social recognition implies, not
just the failure to respond emotionally, but a more general failure to comprehend the significance of social behaviour.

The second component of the triad is the disruption of the normal child's ability, not merely to respond, but to engage in communicative behaviour. This shows itself from the early stages in the failure to engage attention, to point and understand pointing, and, slightly later, in a language deficit characteristic of autism. Indeed, just as the vast majority of autistic individuals are mentally retarded, so the vast majority of autistic children remain essentially mute. However, the range of language abilities, like the extent of mental retardation, is quite marked. The classic autistic language deficit is babbling, in which the child merely makes noises, or echolalia, in which the child merely repeats the words spoken to her.

A detailed study of early autistic language use, which compared the speech patterns of autistic, Down's syndrome and normal children of similar mental ages identified these aspects of autistic language use.

(i) Early pronoun reversal (for example saying "I give you biscuit" when the mother gives the child a biscuit) and failure to understand the difference between questions and statements.

(ii) No mention of cognitive mental states.

(iii) Inability to maintain a conversation: i.e. to extend from the present topic, introduce related topics or collateral information. This ability, which is one of the striking features of normal language development, is absent in even the highest functioning autistics. In the study on very young children this inability showed up in the very low number of "wh-questions" asked by the autistic child. Where the normal child asks "why", "where" and "when" in an attempt to gain further information, and in a way which naturally extends the conversational boundaries, the autistic child does not.

The conclusion the experimenter drew from this study was that "From the onset of intentional communication through the understanding of a formal linguistic system, autistic children demonstrate a fundamental impairment in
their understanding that communication and language exist for the exchange of information or knowledge. ³

This hypothesis is consistent with the linguistic behaviour of high functioning autistics. Even those autistics who eventually master language remain difficult to talk to. Conversation is literal, repetitious and limited to the small number of topics, often aspects of the detail of the physical environment of interest to the autistic person.

The final component of the triad is can also be seen as an extension and precisification of an aspect of autism classified slightly differently by earlier theorists. The play of autistic children is strikingly different from that of normal children. The autistic child will repeat the same physical movement, focus on the same pattern or collect and arrange physical objects obsessively and incessantly. The autistic child does this alone and resents the intrusion of others or any attempt by others to engage in the act of play by introducing a communicative or interactive element. So such normal aspects of childhood play as games of pretence, deception and fantasy, like pretending the broom is a horse and "riding" it have no place in the autistic child's life. Instead the autistic child focuses on details of the concrete physical environment. Wing generalises this aspect of autism and interprets it as a failure of the normal capacity for imagination. Where normal children rapidly develop the ability to pretend while playing and in social contexts, and come to rely on this ability in communicative situations, the autistic child remains trapped by her literal understanding of her environment. In treating lack of imagination as a key element of the autistic triad Wing has added an explanatory or theoretical component to the behavioural classification, since it is partly proposed as an explanation as well as an operational definition.

³(Tager-Flusberg 1993p153)
interpretation of autistic social deficits: lack of social recognition, lack of communication and lack of imagination. Wing’s triad slightly loads the explanatory dice in that it suggests a connection, based in lack of imagination, between the autistic’s involvement with her concrete physical environment and her lack of social involvement. As a result she concludes: "A more likely hypothesis is that a particular area of the brain is concerned with the potential for social recognition, communication and imagination and that the triad occurs when any pathological process leads to its dysfunction." Whatever one thinks of this neurological hypothesis (and as yet there is no neurological evidence which conclusively supports any theoretical interpretation of autism) Wing is surely right to put the emphasis on the abilities, of which imagination is one, which underlie social functioning.

The autobiographical testimony of high functioning autistics gives an insight into their social difficulties. As yet there are a small number of these documents and often they show the influence of the various clinical regimes to which the autistic person almost inevitably becomes exposed (so that the vocabulary in which they express their experience reflects the way in which they were taught the nature of their disorder). Nonetheless they comprise a striking body of anecdotal evidence which confirms that autistics have almost no comprehension of the mental lives of others. "I never could have a friend. I don’t know what to do with people really" is a typical self description. High functioning autistics are aware that they are cut off from others in an important way and may become frustrated and depressed by their inability to understand and overcome their difficulty. Others seem unworried by their lack of understanding but continue to avoid social situations because of the confusion and disorientation which comes with social interaction. In these cases it is the

4 The World Health Organisation’s International Classification of Diseases specifies autism by concrete examples under the following headings:

Qualitative Impairment in Social Recognition
Qualitative Impairment in Verbal and Nonverbal Communication and in imaginative activity
Markedly restricted repertoire of activities and interests

5 (Wing 1988 p103)

6 Reported by Cohen, quoted at (Cohen 1993)
confusion and disorientation in the autistic's world which is resented rather than the inability to communicate with others. In either case, the autistic understands that an important ability which underlies the ability of others to negotiate their social world, is inaccessible to her.

A case is reported by Frith of an intelligent autistic man who had an obsession with detective fiction and toxicological textbooks. Interestingly, and predictably, he could never understand the psychology the detective used to solve the crimes. He could not understand how the detective attributed motives to people and how they connected up with the criminal's actions. In particular he could not understand why someone would lie, so the detective's ability to "see through" the phoney alibis to the real intentions of the murderers was utterly incomprehensible to him. Why this fascination with detective stories if their central feature was inaccessible to him? And why the toxicology books? The suggestion of his psychiatrist is completely consistent with the testimony of adult autistics about how they understand their world". Clearly the fictional detective's ability to impute motive was a source of considerable interest to him. It frightened him that he was unable to do this. He reassured himself that he could solve the crime without psychological insight, by toxicological analysis of the body." 

7(Frith 1989)
One consideration in favour of the theory of mind hypothesis was that it explained, not only the false belief and change of representation tests, but that it generated plausible explanations of the inability to understand what others are thinking and feeling: i.e. to attribute mental states. Can the ToMM hypothesis explain the classic autistic triad of social recognition, communication and imagination? In a weak or minimal sense it is of course true to say that the autistic child lacks a theory of mind in that she is unable to recognise the intentional states of others and to respond and act appropriately. In this sense, however the Rylean would agree that the child lacks practical knowledge of a theory of mind because she has none of the usual dispositions and responses which we classify and explain using mentalistic terminology. However the hypothesis of ToMM is that these abilities depend on modularised capacity to generate M-representations. ToMM thus suggests a hypothesis for the explanation of autism: that the autistic child cannot M-represent. It follows from the ToMM theory that if the ToMM is damaged at any stage of its development then this will be reflected in the inability of the child to master the concepts of intentional psychology. (Section 3.3.) Recall also Wing’s remark "a particular area of the brain is concerned with the potential for social recognition, communication and imagination and ... the triad occurs when any pathological process leads to its dysfunction".

ToMM can be seen as a hypothesis about the nature of the cognitive mechanism whose impairment results in the triad. Not necessarily a hypothesis about the location and structure of the particular area of the brain as Frith misleadingly puts it, but a model of the cognitive processes realised by a modular capacity of the brain. The distinction is important because a mental ability can be modular without being localised to a particular area of the brain, and a model can be true of a mental function without mapping the fine
structure of the brain which supports it. The visual system is an example. Computational models of vision do not propose that there is a single area of the brain which does the computation or that the functions employed by the model have analogical structures in the brain. Nonetheless, vision is a modular capacity in virtue of considerations such as its informational encapsulation and support by dedicated neural circuitry.

How does ToMM fare as a theory of the nature of the autistic disability? The ability of autistic children and adults to recognise and respond to the intentional attitudes of others differs significantly from that of normal people. If this disability is due to a failure of the ToMM module one would expect a significant difference in the performance of normal and autistic children on tests designed to detect the functioning of ToMM. The ToMM hypothesis would be that the autistic child fails to develop or apply an internally represented theory of the relation of mental states to behaviour. A further hypothesis is that theory is domain specific, related only to the generation of intentional explanations.

Consider first the performance of autistic, Down's syndrome and other children on one of the earliest experiments in ToMM research. This experiment is suggested by experiments in the chimpanzee theory of mind. Recall that one interpretation of the chimpanzees' abilities was that they are merely good predictors of typical behavioural sequences and that they predict those sequences without postulating a mental state which intervenes between an agent's perceptions and her actions. Baron-Cohen and his colleagues designed an experiment to test for this possibility in children. They constructed several different sequences represented by four-card story lines. Children were shown the first card in each sequence, then asked to put the rest of the cards in order. The sequences were differentiated according to whether or not they required a

---

8 Because the essence of modularity is informational encapsulation. The information to which the particular model is sensitive confined to that module. Of course the output of one module can be input to another. (Fodor 1988)

9 Another example would be the face-recognition module discussed in chapter 5, whose structure is modelled on Marr's computational model of the visual system.
purely mechanical, a behaviourist or an intentional interpretation in order to be successfully completed. For example one sequence consisted of a person pushing a rock into a river and the resultant splash. To complete this sequence a child requires only causal-mechanical knowledge. Another sequence showed one child taking another's ice cream and the first one crying. This sequence is a behaviourist one because it can be completed by a child who is aware of the typical behaviours of people. It can of course be interpreted in intentional terms as an enviously motivated theft which upsets the other child, but the point of the experiment is that the sequence can be completed by someone who has no concept of these states, but who can remember a typical behavioural sequence. An example of the intentional story is of one child stealing another's teddy bear while the first is distracted, resulting in the first child's look of surprise and dismay. ToMM predicts that only those children, over the age of three and a half, who had full ToMM capacities would be able to perform to the same level on the intentional, behavioural and mechanical sequencing tests. And this prediction was confirmed.

However the main interest in the test was that it was applied not only to normal children but to autistic and Down's syndrome children. The Down's syndrome children did not vary significantly from the normal children, indicating that a general cognitive impairment did not affect their ability to explain situations intentionally. Interestingly, however, the autistic children who also had no trouble with the mechanical and behavioural sequences, performed badly on the intentional sequences.10

This was taken by the early experimenters to show that the autistic children's deficit was related to the ability to make the inferences necessary for intentional interpretations. These results are, however, silent on the nature of those abilities: in particular they tell us nothing about the nature of the representations on which that ability depends.

10 Descriptions of the experiment and details of the sequences tested are in (Baron-Cohen 1986)
4.2.1. The Metarepresentation Hypothesis

One version of ToMM, already discussed in relation to the concept of pretence, is that M-representation is itself a subclass of a generalised ability to form representation of representations: to metarepresent. Remember the diagnosis of the false belief test was that the normal three year old child and the autistic child do not have the ability to understand that others may have a different perspective on their environment than themselves. As a result they are unable to pass the false belief test because they do not have the concept of opaque mental representation. That is to say they are unable to explain others by attributing them with non veridical representations of the environment. Leslie proposed that the development of the ability to pretend was a precursor state in the development of an opaque concept of mental states. Pretence is a state which requires the child, in Leslie's phrase, to "decouple" her mental representations from the environment and to attribute to others similarly decoupled representations.

Consider, however, some recent experiments on false belief and false photographs which build on Zaitchek's findings. Zaitchek trained children in the use of a polaroid camera before conducting the following analogy of the false belief experiment. Children took a photo of a doll wearing a blue dress, while the photo was being developed the dolls dress was changed for a green one. The children were then asked "what colour dress is the doll in the photo?" Zaitchek found that the abilities of three and four year olds did not differ between the false belief and false photo tests, encouraging the hypothesis that the breakthrough represented by passing the false belief test has something to do with the general ability to understand the gap between reality and representation. Hence the metarepresentational hypothesis.

These experiments have been replicated with autistic children with interesting results. Autistic children and adolescents tend to perform at the
three year old level on the false belief test, encouraging the preliminary hypothesis that they lack the metarepresentational ability developed by the normal child at around three and a half. If this is the case then autistic children should perform poorly on the false photograph test. However, recent experiments conducted using photos and drawings show that the same autistic children who perform poorly on an identical false belief test, perform almost infallibly on the false photo and drawing tests.

Perner and Leekham\(^\text{13}\) designed a false belief and false photo test with almost identical structures. In the false belief test children photographed a doll in a blue dress. While the photo was being developed the children were asked to change the doll's dress for a red one and then asked, "What colour is Judy's dress in the photo." Of fifteen autistic children only one got this question wrong. (Charman and Baron-Cohen have replicated this "almost perfect" performance using drawings instead of photographs\(^\text{14}\)). However, nine out of sixteen three year olds and eight out of nineteen four year olds got this question wrong.

The experimenters then conducted a version of the false belief experiment using the same materials. The children and another subject sat in a room looking at the doll wearing a blue dress. The subject then left the room. While the subject was out of the room the children changed the doll's dress for a red one. The child was then asked "what colour dress will the subject think Judy is wearing."\(^\text{15}\) Only four of the fifteen autistic children got this question right, a similar result to the four out of the sixteen three year olds who answered correctly. As would be predicted by ToMM, fourteen of the nineteen four year olds were successful (a higher number than passed the false photo test).

\(^\text{13}\)(Perner 1993)
\(^\text{14}\)(Charman 1992)
\(^\text{15}\)This might seem like an excessively complex way to phrase the question, but Leekham and Perner went to some lengths to establish that this type of syntactic embedding was understood even by the autistic participants. To keep the false belief test exactly parallel to the false photo test they would have liked to ask "what colour is the dress in the photo/picture in the subject's head." But, as they acknowledge, by the time they explained this type of question to the children the question about metarepresentation would have been begged!
These results have been interpreted differently. Perner, focusing on the comparable results of the autistic and the three year old child on the false belief test, argued that the similarity is to be explained in terms of a metarepresentational deficit. The four year old exceeds the three year old and the autistic child on both tests because a generalised metarepresentational ability comes at about the age of three and a half. This hypothesis was advanced in opposition to Leslie’s original proposal that pretence serves as a precursor in the development of the ability to "decouple" mental states from the world. However, the objection to Leslie is principally on technical grounds to do with the computational model of decoupling which he employs in his earlier articles. (Another more straightforward objection is, of course, that if pretence requires metarepresentation, then the child who can pretend should have no trouble with the false belief test. Perner does not make much of this latter objection although it seems to follow naturally if one treats "decoupling" as an instance of metarepresentation.)

However, the main problem with the metarepresentation hypothesis is the "near perfect performance" of the autistic children on the false photo and false map tests. Photos and maps were chosen as devices of representation in order to test the hypothesis that the difference between three and four year olds’ performance on the false belief test might be explained in terms of a general ability to handle representation. However fourteen of the fifteen autistic children pass the false photo test but only four of them pass the false belief test in Perner and Leekham’s experiment. Add to this the fact that autistic children perform significantly better than four year olds on the false photo test and suspicions start to arise as to whether the same ability underlies facility with pictures as underlies mental states.

4.2.2. Metarepresentation and Opacity
Perner diagnoses that suspicion as follows. The photos are indeed representations of external reality, but the autistic child does not conceive of them that way in passing the false belief test: he says the autistic child is a situation theorist, "they simply treat a picture in terms of the situation that is represented in the picture" rather than a theorist who can "understand the picture as a medium which stands in a representational relation to the thing it depicts." Situation theory, although seriously underdescribed here, is perhaps best thought of as a relation of unmediated causation resulting in transparent perception of situations along the lines suggested in Gibsonian psychology. (Transparent is here used in the sense of logical transparency discussed at 1.2. It is not controversial that one can substitute another description of the object perceived into a sentence like "John sees that x is F" without affecting the truth value of the sentence. However the apparent failure of such substitutions to preserve truth value for sentences like "S believes that x is F" generates the problem of opacity). Such a notion is adequate for the false photo task, since all one needs is to compare the situation in the picture to the situation in reality in order to understand that Judy in the picture is wearing a green dress but Judy in the world is wearing a blue dress. According to Perner, however, to pass the false belief test one needs a concept of the situation as represented in the mind of the agent. The situation/representation contrast is thus, for Perner, the same as the transparent/opaque contrast.

There is a contrast between situation theory and opaque representation as Perner describes it, but is it the right way to account for the autistic child's performance on the false photo test? Another interpretation might be that the autistic child understands representations of concrete physical situations better than she understands representations of mental states. Perhaps the autistic child does understand quite well that the drawing is a drawing of a particular situation, and that situation has now changed. So when she compares that

---

16(Perner 1993p129)
17Gopnik also describes the three year old as a situation theorist. The term derives from Gibsonian psychology.
photo with the present situation she is not comparing two situations but a representation of a situation, which has now gone out of date, with the present situation. It may be the case that the autistic child’s mind-blindness and fascination with the inanimate world ensures that representations of that world are far more significant for her. Many autistic children have extraordinary ability to detect patterns and structures as well as tiny details of pictures, scenes and objects with which they are presented. They can often draw these scenes to a far greater level of detail than normal children after only the briefest viewing, often in circumstances in which the average person would remember only the most obvious features. There is no suggestion that when the autistic person does so she does not know that she is producing a picture of a scene. So perhaps the autistic child can understand the representational relation between a picture and a scene, but not the relation between a mental state and its object? This hypothesis would restrict the metarepresentational deficit to mental states only.

This is in fact the route to M-representation. The M-representation theory, as presented by Leslie, recognises that a general meta-representational deficit does not seem to be at the heart of autism, because it appears that autistics can pass analogues of the false belief test for representations of physical objects and other scenarios which do not require intentional interpretation. Thus Leslie proposes that M-representation is a modular capacity devoted to the metarepresentation of intentional states. He also proposes that the representations on which pretence depends are precursors to the development of M-representations.

There seem to be two objections to that proposal. The first is that the autistic inability to pretend could also be seen as part of a general deficit in imagination, which applies equally to the physical and mental worlds. That is to say, some one who has a very limited imaginative capacity is also going to be unable to pretend. Autistics certainly are imaginatively impoverished, the

---

18 Oliver Sacks discusses a classic example in Sacks 1993
quality of their play and their incapacity to conceive of alternatives to their present situation attests to that. The testimony of high functioning autistics also indicates that their inner life is a mainly a reflection of their external circumstances or memories thereof. In fact they find activities which involve the imagination as confusing and disorienting as they find involvement in the social world. The ability to extend imaginatively from their situation does not seem to be part of the autistic mental repertoire. If this imaginative deficit is prior to, or subsumes, the pretence deficit, then inability to pretend will not be a failure of the precursor stage of metarepresentation, but an inability to imagine a pretend scenario.\textsuperscript{19}

Metarepresentation is only one possible answer to the question "What, if anything, does the problem of opacity tell us about the sort of thinking we use when understanding other people? " Philosophers and psychologists who favour the M-representation theory of mind have been heartened by the fact that many logicians have treated intentional states as a relation between a thinker, an object and a proposition and have accorded the represented proposition the ultimate privilege of quantifying over it.\textsuperscript{20} But, as I have emphasised, this is a matter of philosophical convenience and the economical

\textsuperscript{19}There are, of course theories of intentional psychology which would make \textit{imagination} the key to intentional understanding. That is, they explain the ability to attribute mental states in terms of an ability to \textit{imagine} "what it is like" from another's perspective. Such theories bypass ToMM entirely since they dispense with the need for inference, based in a theory of mental states, in order to understand what another person is thinking or feeling. Of course, the determined ToMM theorist can always maintain that the theory of propositional codifies of practical knowledge realised by an imaginative capacity: but that would be an attempt, not to save the phenomena, but the ToMM theory from the phenomena. See the symposium on Simulation theory in \textit{Behavioural and Brain Sciences} (1993) 16. Someone who takes simulation theory seriously and has applied it to the Autism case is Currie. (Currie Forthcoming)Such an approach treats the failure of imaginative capacity as the basic cause of the autistic triad. I am not hostile to such an approach, since, like Ryle, I am sure that a good part of deliberation and of interpersonal understanding may involve imaginative or empathetic rehearsal of situations and emotions. However my account would treat an imaginative capacity as one of several capacities which may need to be integrated to construct a context for intentional understanding. That is why I do not go into the debate between CTM theorists and Simulation theorists in any detail. Vindicating simulation theory as a \textit{general theory} of propositional attitude psychology depends on how well it survives the attacks of the like of Stich. On Simulation theory see also Blackburn(Blackburn 1992)

\textsuperscript{20}I think Kaplan makes the breakthrough here in his discussions of Quine on intensionality. (Kaplan 1985) See also Burge (Burge 1977a) and (Burge 1977) The most explicit recent statement I know of one version of the position is taken in Perry's recent work, where he takes himself to be quantifying over something like sentences in a language of thought: brain states with semantic properties (Perry 1989,Perry 1990,Perry 1991)
rendering of logical form. Intentional states as they figure in philosophical logic are actually placeholders for the psychological property which underlies the ability to produce, interpret and explain intentional behaviour. Logic alone does not tell us whether or not intentional states are attributed via M-representations. When a logician quantifies over an intentional state she is not legislating about the nature of intentional states (except to say that they exist) but the nature of sentences which talk about them. Intentional states may be sentences in inner mentalese, weightings in a connectionist network, patterns of dispositions, second order functional states, mental models, simulations or another candidate to be delivered by a completed psychological theory. Logic does not tell us which, if any, of these candidates actually explains intentional action.

The ToMM theorist assumes that the only things to which sentences about intentional states could refer are M-representations. In stating a person's belief we are not just concerned about how that person's mind looks, but how that person thinks the world is. And to understand that belief is about how the world is a representational view of mind is necessary (my italics) But there is no a priori necessary connection between the logical form of sentences (which contain embedded propositions) used by a peripheral theory which attributes states of mind, and the structure of the mental representations which realise those attributions. (Depending on one's views about semantics and psychology there may be a necessary a posteriori connection. But even if this is the case, the metarepresentation view has the status of an empirical hypothesis about the structure of second order mental states, and stands or falls on the results of empirical psychology, not of logic.) Leslie and Thaiss, commenting on Perner's proposal, make a similar point. They note that Perner's metarepresentational hypothesis draws support from a treatment of opacity, suggested by Perry and Barwise in Situation Semantics, which involves

21(Perner 1993p129)
quantifying over a representational relation. However, they point out that, even if Perry and Barwise are right about the semantics of sentences which express propositional attitude ascriptions, "there are no clear grounds for supposing that the above relational analysis is only possible if the child projects a representational medium into the mind of Sally."22 (my italics)

4.2.3. A Dispositional Interpretation of the Evidence for Metarepresentation.

I have argued that Ryle’s analysis of mental states as multi tracked dispositions captures more of the intuitive meaning of mental state terms than the computational model embraced by the ToMM theorist. On a Rylean view, the difference between the autistic child, the three year old, and the four year old is in their differential sensitivity to the dispositions of their fellows, a sensitivity which is manifested in their own dispositional profile. Our reactions to the autistic case tend to bear this out. The autistic child is remarkable for the deviance of her dispositional profile and her inability to coordinate her responses and actions to those of others.

However, even if Ryle is right about this, the metarepresentational theorist might claim that the differential dispositions of the autistic and the normal child can be satisfactorily explained in terms of a difference in metarepresentational ability acquired at the age of three and a half. So, even if Ryle wins the argument over intuitive semantics, the substantial metaphysical and epistemological issues are resolved in favour of the metarepresentational theorist.

A Rylean cannot rest here, because on his view the semantic issue is intimately connected to the other two. According to Ryle what we know when we know someone else believes the ice is thin is that they are disposed to skate warily, to dwell in imagination on possible dangers, to warn other skaters etc. To paraphrase Ryle we credit them with a host of executive and imaginative

22(Leslie 1992p243)
abilities. In crediting them with those abilities we rely on a host of our own executive and imaginative abilities, and there is no \textit{a priori} reason to think that these abilities are all sustained by the same modular capacity of the brain.

Consider the difference between the three year old and the autistic child. Both of them typically fail the false belief test. Yet the three year old is a demonstrably superior intentional psychologist to the autistic child. She already has most of the abilities needed to understand action in context. She lacks only the ability to understand contexts where one needs to infer that the content of another's beliefs will not be a straightforward mirror of the situation as perceptually presented. But even here, as when she fails the false belief test, it seems that she fails it in a different way to the autistic child. The three year old views the scenario presented in the false belief test as one to be interpreted in a psychological, communicative context. She gives a false interpretation because her repertoire of psychological concepts is narrower than that of the four year old. There is no evidence that the autistic child constructs the context in the same way, as one whose interpretation requires information about the psychological states of another person. So to describe the three year old and the autistic child as both being situation theorists is misleading, because the repertoire of interpretative skills they bring to bear on the situation are so different. The three year old treats others' behaviour as laden with intentional meaning, even if she is wrong about the exact meaning, whereas the autistic child seems to ignore the intentions and motivations of others when it comes to explaining their behaviour. Thus they both fail to detect what the other person is thinking in the False Belief scenario, but for different reasons.

Leslie is right to point out that the normal child is constantly involved in activities which involve communicative interaction and reading meaning into others' behaviour. What are described by ToMM as precursive abilities to full ToMM all take place within a communicative context. Significantly, the autistic child does not exhibit these precursive abilities. Autistic children do not engage in early imitation, gaze and attention monitoring, proto-declarative pointing,
the use of language which talks about cognitive states, spontaneous pretend play, or interactive sociability in general. All these activities can be seen as involving the attribution of mental states at increasing levels of sophistication. The ToMM characterisation allows that these abilities can be unified in terms of the development by the child of a cognitive theory, which employs successive theoretical models of her social environment. The metarepresentation hypothesis advanced by Perner is that the last stage in this process, the final model represented by ToMM, is a theory of mind which treats mental states as attitudes to propositions.

This is an elegant hypothesis, since it unifies a range of phenomena: the differential abilities of the normal and autistic child and the false belief test. It also explains the inability of the autistic child to develop a metarepresentational capacity, in terms of the failure of one of the precursive states, perhaps as early as the inability to imitate.

The Rylean hypothesis is less elegant but also less demanding. The Rylean would say that interpersonal understanding involves a variety of abilities which add up to practical knowledge of a theory of mind, but which are not unified via the internal representation of that theory. Really this is not so different from the claim of commonsense functionalism described at 1.6. The claim of the commonsense functionalist is that behaviour is caused by the states of the agent’s brain which sustain the belief-desire network we invoke in holistic interpretation of the agent’s behaviour. The commonsense functionalist, in contrast to the psycho or empirical functionalist is not committed to a specific hypothesis about the way in which intentional states are realised in the brain of the agent.23

23Historically, functionalism is a move away from a version of Ryleanism. It adds the claim that it is not just the dispositions which are important to the attribution of mental states, but that those dispositions have a categorical base in the brain of the agent. The commonsense functionalist establishes this by conceptual analysis. See eg (Armstrong 1968) and for an elegant recent statement of the position (Lewis 1994) or (Jackson 1993). The psychofunctionalist, who is Churchland’s target, adds to commonsense analysis a theory that the categorical base, the brain of the agent, represents the structure of intentional psychology in symbolic or linguistic form.
Note that all I am arguing for here is the claim that the abilities the child displays do warrant the attribution of any cognitive abilities stronger than practical knowledge of the truths of intentional psychology, realised by a variety of knowhows and capacities. I have not argued that the information processing on which the child's ability to interpret behaviour depends is not encapsulated, so I have made no claim about modularity per se. However it is important to notice that the ToMM theorists I discuss never mount a sustained discussion of the issue of informational encapsulation, tending to assume that a domain specific ability, like intentional interpretation, which can be unified by some fundamental principles depends on the functioning of a module which represents those principles.

At this point the ToMM theorist may claim that neutrality on the issue of the nature of the mechanisms which which sustain a dispositional profile and emphasis is a weakness. She might agree with Ryle that what we do when we explain someone else's behaviour is to systematise their dispositions, but she adds the hypothesis that this integration of dispositions depends on the functioning of a modular capacity of the mind with distinctive properties. Thus, even though she might be prepared to concede that the young child does not engage in explicit theorising, she provides an empirical hypothesis about the way the mind actually works.

My response to this point is to say that the advantage of the neutrality thesis is that it does give us a better understanding of the actual mental representations involved in intentional understanding. ToMM is a substantial but false empirical hypothesis whose failings are masked by an interpretation of the evidence of developmental psychology which treats it as a modular device for generating theories which represent propositions of a certain form: namely Metarepresentations. This foreclosing of possibilities is partly the result of a dialectic between behaviourists and mentalists in the philosophy of mind which structures the way experiments are constructed and interpreted.
One reason the conception of intentional psychology as a species of theory construction has been so attractive and powerful in many fields is that it is not a behaviourist thesis. Behaviourism requires us to understand mental state terms as referring to patterns of behaviour alone, leaving out any reference to the internal processing, subjective experience and the conscious mental activity which enters into the production of behaviour. This is objectionable for at least two reasons: the first is that, for humans at least, thought and planning seem to be what separates intelligent action from mere bodily movement, and it is the production of intelligent action adjusted to the physical and social context, which psychology seeks to explain. The second is that it is implausible to think that we adjust our behaviour to others purely on the basis of induction from observed regularities in past behaviour and predictions about future behaviour without making any inferences about the underlying processes which unify those patterns. In fact it is claimed that intelligent behaviour cannot be accounted for in this way.

ToMM is attractive precisely because it goes beyond the limitations of observable behaviour. So the dialogue between ToMM theorists of mind and others often looks as if the only alternatives are bare behaviourism and a ToMM. Since bare behaviourism cannot account for the range of abilities which attract psychological explanations, we posit an internally represented theory which can unify those abilities and interpret the behaviour mentalistically. This can be seen in the literature on both the chimpanzee theory of mind and the child's theory of mind. Recall that Heyes and those who wished to provide a conservative interpretation of the chimpanzee's abilities did so by showing how they could be accounted for in terms of their ability to recognise typical behavioural sequences without the postulation of any "inner" states. Similarly many theorists who discard the hypothesis of an internally represented theory
of a certain domain ask if the specific ability can be reinterpreted behaviouristically.

Baron-Cohen's initial experiment is structured by the assumption that the alternatives available to explain the child's abilities are exhausted by behaviourism or an internally represented theory of mind. He set up a series of sequences which require mechanical, behavioural or intentional interpretations in order to be successfully completed. The idea was that the difference between a purely behaviouristic and an intentional interpretation was the attribution of intentions to the actors in the sequence. This is surely correct, but must we assume that the attribution of intention requires an inference from behavioural evidence to an underlying state of mind? Could it not be the case that attributing intentions is a more sophisticated way of classifying behaviour but one which does not depend on theoretical inference of the nature of underlying causes from evidence which consists of uninterpreted behaviour? Leslie and Thaiss make a similar suggestion when they say that "the intuitive theory of mind is not concerned with this question [a scientific attempt to explain what sort of things propositional attitudes really are] but with capturing how meaning enters into the causation of behaviour." 24

M-representation theorists do not disentangle the two questions. They assume that in order to read meaning into behaviour the child needs a theory in order to infer the content of the underlying representations, which are the bearers of meaning. A classic instance of this dilemma, behaviourism or hypothetico-deductive postulation of internal representations, is provided by Sodian and Frith in their work on deception in normal and autistic children. Sodian and Frith conclude, reluctantly, that in the case of autistic children who master deception and pass variations of the false belief test that further investigation is needed "to establish if there is evidence for a theory of mind in their everyday behaviour or whether they have developed relatively sophisticated strategies to compensate for their deficit...It is tantalizing that

24 Leslie 1992 p245
clear cut evidence for the employment of mere behavioural routines versus deeply understood concepts remains elusive."

I have already discussed the difficulty of deciding whether or not deception involves the exploitation of a theory of mind in order to create a false belief. In the case of the chimpanzees I argued that it does not, but that does not mean the chimpanzees do not understand deception. Their practical knowledge of the concept in fact exceeds that of infant humans. However, and this is a point which arose in discussion of the chimpanzees, practical knowledge is not the same thing as behaviourism. One might be a good predictor of behavioural sequences precisely because one has practical knowledge of the intentional states needed to interpret those sequences.

At the very least, the theory of mind experiments show that children are never behaviourists. They are innately equipped to attend and respond to some expressions and gestures rather than others. As they grow older they monitor others' attention, produce proto-declaratives and play games of hide and seek. In all these activities the child is recognising and responding to the mental states of others with mental states of its own. The ToMM theorist of course does not propose that these abilities depend on fully fledged ToMM, she treats them as precursors in the development of the full theory. So the ToMM theorist grants that there are a range of social abilities, such as those possessed by three year olds, which do not depend on the projection of representations into the heads of others via a hypothetico deductive theory. The important point is that the ToMM experiments establish that the evidence with which the baby is confronted is evidence of other people's attitudes towards her, not bare behaviour which requires theoretical interpretation before it can be understood as communicative or meaningful.

The ToMM theorist can grant this, but she will claim that the four year old, if not the baby, needs Gopnik's "coherent, abstract account of the mind" in order to negotiate her social world. She cannot rely on a mixed bag of innate

---

25(Sodian 1993pp174-175)
recognitional capacities, practical knowledge of typical actions, emotional responses, linguistic and other cognitive knowledge in order to understand other people’s actions in the social context. Of course she may, on occasion, rely on these capacities, but she would not be able to negotiate the complexities of the social world if these disparate abilities were not unified by a conception of the mind which classified them as attitudes to represented propositions, and projected representations with a specific structure, into the minds of others in order to explain their actions.

So the social deficits of the autistic child can be claimed to support either a Rylean or the ToMM analysis of social understanding. So far, however the field has been monopolised by ToMM in virtue of two main claims. The first is that in explaining ToMM we have a choice between embracing ToMM or an account which treats the child as a behaviourist. The second is that the child’s global social understanding is best explained in terms of an internal theory which unifies understanding of mental states by treating them as attitudes to internally represented propositions.

It is true to say that the Rylean will want to explain the difference between the autistic, the normal three year old, and the normal four year old in terms of their abilities, manifest as dispositions, to interpret behaviour. However that is not to embrace a "bare behaviourism”. The Rylean, just like the ToMM theorist, can acknowledge that the child understands the meaning of much of the behaviour with which she is confronted, in virtue of a battery of responses and practical capacities. On the second point, the Rylean may dispute the claim that intentional understanding depends on the deployment of an overarching theory.

Note that this version of the Rylean hypothesis is quite consistent with the following version of child development:
interaction. Eventually it can interpret and predict the actions of others, even in complex social contexts, and correctly use and apply propositional attitude terminology. The language of propositional attitude psychology treats all mental states as attitudes to propositions (without making any claim that such attitudes are the internal tokening of representations of the content of those propositions) and so it gives the child a unifying vocabulary with which to treat all the abilities relevant to social understanding. So the mature child can employ this vocabulary to work out "in its mind" what other people might do in counterfactual situations and to understand that others may often deliberate in the same way. Hence when it refers to the deliberations of others, or their reasons for acting, the child can cite beliefs or desires, understood as acts of deliberation, the site of which is the other person's mind.

This looks very like what the child actually does, and it is consistent with the developmental story told by ToMM theory, but it does not imply that the child projects representations into the minds of others in order to explain their behaviour. ToMM is only vindicated by this account if the ability to deliberate using propositional attitude concepts and to use propositional attitude vocabulary depends on the internal representation of a theory of propositional attitude psychology (or its mentalese translation). Tager-Flusberg makes an interesting comment on this point in her discussion of the minority of autistic children who pass the false belief test. As a ToMM theorist she needs to explain how it is that such children can do so without an internally represented theory of propositional attitude psychology. She offers an explanation very close to the one Ryle and Sellars give of normal intentional understanding. "It may even be the case that for some children, experience in using these cognitive state terms and the syntax of embeddings more generally are instrumental in promoting the conceptual developments that underlie the acquisition of a belief-based psychology".26

26(Tager Flusberg 1993)
Recall the musician example in chapter two. A musician who learns to read music may acquire a representative medium in which to articulate, to herself, introspectively, and to others, her ability to play music. However that is not the same thing as explaining her musical ability in terms of inner representation of a musical score, or mentalese version thereof. Of course if one could only explain the ability to read music in terms of a prior act of inner musical mentalese then one would be justified in inferring the presence of musical mentalese from the ability to read music. But there is no necessary connection between the applicability, or even the conscious employment, of a peripheral theory of an activity and the internal representation of that theory.

The same may be true of children who are able to attribute beliefs and desires. The language of propositional attitude psychology gives them a medium in which to reason about and report on behaviour, including their own deliberative activities, but their ability to use this medium does not imply that the abilities reasoned about are analogous in structure to the reporting medium.

This is where the analogy with physics or chemistry which figures in the neo-Quinean account of intentional explanation misleads the ToMM theorist. Physics and chemistry are theories which do state hypotheses about the structure of underlying processes. However, not all theories attempt to specify the structure of the causally efficacious entities and processes which underlie the dispositions they systematise. Many ToMM theorists seem not to distinguish the structure of the theory of an activity from the nature of the psychological processes on which the activity depends. This is apparent when we look in more detail at the explanation they give of autistic social deficits.

4.4. The Autistic Dyad: Language and Social Recognition in Context

Humans may be capable of controlling more than one technique for performing nondemonstrative inference. A scientist self-consciously applying explicit standards of
confirmation to each piece of available evidence may well be using a quite different system from those we all use — scientists included — in making spontaneous, instantaneous, unconscious inferences about the movements of other vehicles while driving a car, about what some appetising food might taste like, or about a speaker’s communicative intention.\(^{27}\)

In chapter two I discussed relevance. A primary aspect of relevance is the ability to recognise the communicator’s informative intention: i.e. that she is engaged in an act of ostensive communication intended to convey certain information. Relevance is a pragmatic phenomenon because deriving that information depends on the ability to process background information not conveyed by the propositional content of a communicative act. This background information is a fluctuating set of mutual presuppositions which participants in the communicative enterprise must jointly hold in place in order to allow the communication of information: the context of communication. Sperber and Wilson describe the context as a psychological construct, a subset of the hearer’s assumptions about the world. It is these assumptions, of course rather than the actual state of the world, that affect the interpretation of an utterance. A context in this sense is not limited to information about the immediate physical environment or the immediately preceding utterances: expectations about the future, scientific hypotheses or religious beliefs, anecdotal memories, general cultural assumptions, beliefs about the mental state of the speaker, may all play a role in the interpretation.\(^{28}\)

Those interested in the psychological processes underlying pragmatic communication have to explain, as Sperber and Wilson show, not just how the speaker’s meaning can be recovered from the context and the proposition expressed, but how the utterance creates contextual effects: i.e. how the hearer comes to generate or modify the context as a result of the communicative act.

\(^{27}\)(Sperber 1994 p67)
\(^{28}\)(Sperber 1994p15)
It is here that the autistic language deficit is interesting because the autistic language deficit is a pragmatic rather than a semantic one. High functioning autistics understand the lexical meaning of what is said or written well enough, but it creates no contextual effects for them. As a result, unless the conversation is strictly literal and/or the relevant contextual assumptions are also described literally and in detail, they can never recover the speaker's meaning because they cannot provide a context.

Since the context is a psychological construct (Fauconnier calls it a "mental space" whose dimensions fluctuate as the conversation proceeds\(^\text{29}\)) an autistic person is clearly at a disadvantage in understanding what must be presumed or presupposed in order to sustain communication.Autistic children as young as nine months demonstrably lack this ability. The normal child starts to use proto declaratives at about this age and to respond to their use by others. This is one of the first clear instances in which the autistic child is unable to construct an appropriate context, which would require the concept of joint attention and the concept of whatever mental state, (amusement, surprise etc) accompanies the other's pointing. In the discussion of pretence I noted that the normal child's development of mentalising abilities takes place primarily within a communicative context. Proto declaratives are merely an extension of this phenomenon of communicative engagement which is present from earliest infancy in games of imitation, joint attention and, later, proto imperatives. In the very early stages, the context created may be no more than the sharing of affective states and the communication of emotions of love and caring. Attention monitoring and proto imperatives set the scene for the communication, by gesture, of information about the environment and the child's attitude towards it. Interestingly, autistic children do not seem to monitor gaze,\(^\text{30}\) and although they do use proto imperatives they do not seem to use them communicatively, in the sense that they are designed to produce a response in the minds of their audience. One characteristic feature of autism is

\(^{29}\text{Fauconnier 1985}\)
\(^{30}\text{Baron-Cohen forthcoming}\)
the taking of a parent by the hand and dragging them to a desired object rather than pointing to it. Even from the early stages, then, autistic children are communicatively impaired.

As the normal child acquires language she starts to exhibit typical features of conversation which require a far more complex construction of context than that required by proto declarative communication. The most significant feature of the developing child's language ability is the ability to extend appropriately and add new, relevant information to the topic of conversation. The autistic child never acquires this ability.

Even from the early stages, the autistic child is unable to construct an appropriate context, even for gestures. Is this evidence that the autistic child lacks a "theory of mind" perhaps as a result of damage to the modular ability which develops through the typical precursor stages of imitation, joint attention, proto imperatives and declaratives? Tager Flusberg thinks so: "The specialised and multifaceted forms and functions of human language appear to have been designed for communicating with other people about mental states. One of the primary forms of language, to serve as a major source of knowledge, is impaired in autistic children even in the prelinguistic period. It is this impairment which links deficits in joint attention, later problems with communication, and the understanding of belief."

However this summary is misleading. What is the evidence that the specialised forms and functions of human language appear to have been designed for communicating with other people about mental states? Language is more often seen as a medium for communication of information in general. Of course, to understand what is communicated, to derive the intended information, the hearer needs to know a lot about the speaker's mental states in order to construct an appropriate context. There is evidence that children develop their abilities to do this, and indeed cannot do this without, immersion in a series of progressively more complex communicative environments. As the

31 (Tager Flusberg 1993)
environments become more complex the range of abilities needed to construct a context also expands. It is not surprising that a child who can pass the false belief test and attribute opaque mental states is able to construct appropriate contexts for the vast range of communicative acts which are not strictly literal, or for which the proposition expressed is insufficient to communicate the speaker's meaning. And it is not surprising that children who are already competent with games of pretence and deception and who have a lively fantasy life centred around communicative engagement are better equipped to attribute the appropriate intention in nonliteral contexts.

So the construction of a context depends on the ability to attribute mental states, including belief, where the understanding of the speaker's belief may add important contextual information, and children acquire the ability to do this. This is only evidence of a "Theory of Mind" in the sense under discussion if the child's pragmatic abilities can be shown to depend on the attribution to others of mental states, via a theory which projects representations into the minds of others. The primary evidence that children employ such a theory is, of course, the false belief test. The secondary evidence is the use by children of mentalistic vocabulary which talks about "in his mind", "in her head", "in her imagination" etc.

However, the secondary evidence can be explained in the Rylean way as the acquisition of a vocabulary in which children can articulate their deliberative processes and hypothesise about the deliberations of others. The mastery of a folk vocabulary of intentional states, which can be applied to deliberative acts does not, however, seem significant once one distinguishes the mastery of a peripheral theory from the psychological processes which support its use. A child who sees the dog jumping up and down excitedly when the fridge is opened might say "Lassie knows it's dinnertime". Should we deduce that the child is stating a hypothesis about the nature of a representation entertained in the dog's brain? It is more likely that she is just reading meaning into the dog's behaviour, employing a familiar vocabulary to do so.
Even if we accepted the claim of the ToMM theorist that the false belief test was evidence of the ability to attribute beliefs, conceived of as internal representations, this would not prove that the child’s pragmatic abilities depended on this type of mental state attribution. Constructing a context requires more than the attribution of beliefs, true or false. Sensitivity to emotions, facial expressions, body language, affect, patterns of behaviour, all contribute to recognising the relevant psychological states of the other person and are all relevant to determining the psychological aspects of context.
Chapter Five. Cognition in Context

Precis

The chapter explains mindreading in terms of the construction of a context for the interpretation of intentional activity via the employment of somatic markers.

5.1. Elements of context. The variety of information required to interpret activity. The distinction between theoretical and processing economy. Theoretical economy is elegance or parsimony of explanation, processing economy is actual time taken or energy expended by the brain on a certain task.

5.2. Emotion and Cognition. The emotional and cognitive aspects of context construction normally go together. The example given is the ability to recognise familiar faces. Prosopagnosia and Capgras' syndrome are bizarre failures of the face recognition system best explained as an inability to integrate emotional responses with visual information.

5.2.1. The Role of Emotion in Context Construction. The argument extended from face recognition to social cognition.

5.3. Emotion and Somatic Marking. Damasio's account of emotion as the representation of body state. Certain representations on which social cognition relies are somatically marked, associated with typical primary or secondary emotions.

5.3.1. Evidence for Somatic Marking. The neurological and experimental evidence for the somatic marker hypothesis.

5.4. Context Construction and Somatic Marking: the Autistic Deficits. Autism explained as an inability to construct a context due to the failure to somatically mark psychological elements of intentional situations. Patterns of intentional activity thus never become salient for autistics.

5.5. Philosophical connections. The illusions of expertise explained as a result of somatic marking.

I have claimed that mindreading can be explained without postulating a single module devoted to the production of M-representations. The ToMM theorist may,
quite reasonably, respond" I have given an interpretation of the evidence which unifies a range of disparate phenomena. It is all very well to keep talking about dispositions which are integrated in the way Ryle suggests, but what mental processes underlie the integration of these dispositions if the ToMM account is wrong? If the neutrality thesis commits us to saying nothing about these processes it marks no theoretical advance". This chapter is an answer to this charge, an attempt to show how the neutrality thesis points us toward a different account of the relevant representations.

5.1. Elements of Context

The ToMM theorist herself allows that we recognise many mental states without subsuming them under theoretical concepts, because she regards abilities like responding to affection, monitoring attention, and imitation as precursive of, not identical to, the theoretical abilities she postulates. But is there any evidence that, as we grow older, we cease responding to emotions and expressions in this way and instead subsume them under an internally represented theory of mental states? Of course as we grow older we can classify the interrelation of emotion, affect, perception, expectation etc. using propositional attitude vocabulary, and if we had to give an account of a context we would do so using that vocabulary. It does not follow from the fact that we can attribute propositional attitudes and construct a context, that we conceive of all the psychological components of the context as M-representations. I shall argue that we continue to respond to emotion and to facial expression, for example, in the same direct unmediated way we did as infants.

Contexts often involve attributing an imaginative state, memory or a fantasy which it is natural to understand as an inner state. However, that same context may also involve sensitivity to others' dispositions, as per the chimpanzee or two
year old, and recognition of emotional expressions as well as perceptual recognition of the state of the environment as perceived or, perhaps falsely remembered, by the other person. This would be the exploitation of a mixed bag of recognitional capacities and abilities, manifest in one's disposition to respond appropriately.

When Sperber and Wilson say that context is a psychological construct they mean that it requires certain psychological abilities, including the ability to recognise others' mental states. The abilities implicated are those which would produce the maximum comprehension for the minimum expenditure of psychological processing. Propositional attitude psychology is a theoretically economical way of explaining the relationship among mental states and constructing a context which enables meaning to be read into behaviour. However, theoretical economy is not processing economy: processing economy is something like the number and complexity of computations involved, or the amount of memory, short term memory and attention required by the brain to do a certain task. Why would a system internally represent a set of propositions qua theory if it could achieve the same degree of communicative or behavioural coordination by representing the information required via the rest of its cognitive repertoire: linguistic representation, disposition, innate capacities for facial and expressive recognition, emotional and affective responses? Imagine the process of understanding that someone is angry, often an important contextual feature, but one seldom explicitly conveyed. Any, or all, of the speaker's tone of voice, aggressive posture, stalking around the room, facial expression, or the situation, may all make this manifest. If this is the case then one does not need to project an inner state of anger into the other person's mind in order to tell that they are angry.

One thing we can be sure of is that autistic people cannot construct an appropriate context. This inability to construct a context and engage with others is apparent from almost the beginning of the autistic life in their failure to imitate,
monitor attention, play or point in ways which suggest involvement with the other. ToMM theory suggests that the abilities which the autistic lacks are part of the same module, devoted to interpersonal understanding, which eventually becomes a theory of other minds. In contrast, I propose that the abilities which the autistic lacks, even as a young child, are typical of the range of abilities needed to construct a context and on which we all rely, but which need not depend on the a ToMM module.

5.2. Emotion and Cognition

ToMM treats the acquisition of propositional attitude psychology as a cognitive phenomenon, an instance of intellectual rather than practical mastery of propositional attitude psychology. I suggest that the unifying aspect of the multiple abilities on which folk psychology depends may not be found in the cognitive domain. Recall that some twenty per cent of autistics eventually pass the false belief test, and that another defining criterion of autism is "islets of ability". Nonetheless the language deficit and the deficit in social recognition still persist in these cognitively normal or exceptional (in a limited domain) people. And in these cases the deficit noted by Kanner is still striking; the lack of affect and the inability to detect and respond to the thoughts and feelings of others. It may be this lack of affect which is the unifying phenomenon among the series of deficits which ToMM treats as precursive of ToMM.

That there is a link between the affective and the cognitive aspects of interpersonal understanding is demonstrated by the bizarre behaviour of those in whom the link suddenly breaks. The most striking examples are from the psychological literature to do with face recognition. Prosopagnosia is a condition in which people lose the ability to identify the most familiar faces: friends, family and even the patient's own face in a mirror. Such a person, however, can still
recognise the person, by gesture, voice or other characteristics, only the face remains unfamiliar. Most often prosopagnosia is seen in patients who have suffered damage to bilateral areas of the occipital cortex after a stroke. The explanation proffered until recently was that the stroke damaged one component of a face recognition module which operates in the following way. When a person is confronted with a face the visual information is encoded. That information is then stored in a face recognition node which contains information about expressions and the physical features of the person. When the face is seen again it is checked against that node for recognition. That module is connected to another which stores other sensory information about the person: their voice, gait, body shape etc: the personal identity node. The total input from each node is linked together and connected to a name generator.1

The explanation generated by this model is that a stroke, epileptic attack or other incident damages the face recognition units or the processors which feed into them, but leaves the name generator and the node which stores other semantic information intact. This is why the person with prosopagnosia is able to recognise familiars by relying on other cues. (In one case the size of a husband’s feet.)

Prosopagnosia, as described, is a purely cognitive phenomenon, a defect in the system which stores and retrieves information. So people with prosopagnosia still feel the same emotional and affective attitudes towards the people they can no longer recognise. This is what makes the condition so disturbing for those who have it. The faces of their friends and relations may, depending on which aspect of the module is damaged, appear blurred or distorted. Alternatively, if the connection to the personal identity node is damaged, prosopagnosics may see the face all right but still be unable to recognise the person. They know the face, but not to whom it belongs. So the sense of personal engagement is still there but not attached to its normal anchor of facial recognition.

1 For a fuller description and diagrammatic representation see (Ellis 1990)
The prosopagnosic experiences just how strange it is to interact with people when this recognitional capacity disintegrates. Even stranger is the case of Capgras syndrome. Capgras syndrome was first reported in 1923 as "l'illusion des sosies" experienced by a 53 year old French woman. She believed that people she knew had been replaced by identical doubles. This belief, accompanied by paranoia, is the most typical feature of the syndrome, although Capgras sufferers tend to suffer from a generalised depersonalisation-derealisation. They may think everything looks strange and synthetic, that faces look waxen or objects look painted. However the central feature of the disorder is the illusion that people close to the sufferer have been replaced by identical doubles. A typical case is of a 60 year old woman who believed that "There's been someone like my son's double which isn't my son.... I can tell because my son is different...."2

Capgras sufferers were tested for their ability to recognise faces, with an interesting result. They typically did not misidentify the faces of strangers or people they did not know well, but could not identify the faces of relatives or friends. The hypothesis naturally suggests itself that Capgras syndrome is somehow bound up with damage to the face recognition system. This is encouraged by evidence that most sufferers have received some sort of organic brain damage to the occipito temporal lobe. It has sometimes been successfully treated with psychotropic drugs whose effects are directed to this region of the brain.

However, for this hypothesis to be sustained, the model of face recognition described in the discussion of prosopagnosia needs to be modified. As it stands, it is inadequate to account for Capgras symptoms. The Capgras sufferer does not seem to have a defect in the information processing device. All the information normally received through the visual system, and the semantic information stored at the personal identity node is unaffected. This is why, in the classical cases, she calls the

2(Young 1993 p697)
misrecognised person a double or replica. The second point to note is that the
defect is more pronounced in the case of familiar people. This seems to indicate
that there is more to recognising familiars than correctly identifying them via the
information carried by the visual and personal identity systems.

When Wittgenstein said "My attitude toward him is an attitude toward a
soul. I am not of the opinion that he has a soul". he was offering a response to
scepticism about other minds. He was saying that treating other people as
intentional agents and engaging with them emotionally comes naturally to us. We
do not see bare behaviour, make an inference as to the likely cause and plan a
response. Rather, emotional response is part of the act of recognising that someone
else is unhappy or jubilant. In the words of his more recent followers, the
emotional response to another’s pain is part of the content of the judgement that
she is in pain. If we sometimes engage in hypothetico-deductive reasoning about
mental states, as in psychoanalysis, these are specialised and rare instances. The
normal case is to directly see and respond to others via a range of direct emotional
and recognitional capacities.

Wittgenstein was making an a priori claim about the concepts involved in
intentional understanding: that having the emotional response is part of the act of
understanding, advancing this claim against reductive scientific approaches to
intentional understanding such as pure behaviourism or versions of Cartesianism3
which treat intentional explanation as a purely cognitive phenomenon. Whatever
one feels about the a priori nature of such claims, the cases of Capgras syndrome
and prosopagnosia are easily explained on the Wittgensteinian model. Instead of
the usual affective response to family and friends the Capgras sufferer feels fear
and paranoia. In her at least, the normal direct connection between visual
recognition and emotional response has disintegrated. The cognitive aspect of

3 Although Descartes himself might be exempted on the ground that he explicitly said that every
judgement involves the will.
recognition remains, but she feels none of the emotions which normally accompany the recognition of a familiar person.

Some findings about covert recognition by prosopagnosics support this hypothesis. In prosopagnosia the visual information system is impaired. Nonetheless, prosopagnosics still correctly identify people even though their faces may be blurred or simply unrecognizable. In many cases this is because they rely on information stored at the personal identity node about other physical or historical characteristics. There is, however, another route to recognition available for the prosopagnosic, even if the personal identity node is also disrupted.

The so called "Guilty Knowledge Test" is used by American police forces to test whether or not people are lying about their knowledge of people or events. It measures automatic responses, such as skin conductance response which are beyond the intentional control of the suspect. Bauer applied this test to a profoundly prosopagnosic person. When this person was shown pictures of familiar and unfamiliar faces he naturally said he could not discriminate between them. However his skin conductance responses tracked the familiar faces. That is to say that his skin conductance response increased in only those cases where he was confronted by a familiar person. This type of experiment is not unlike experiments in which clinically blind people nevertheless display discriminative capacities which track fluctuations in light intensity or the location of objects. A proposed explanation of blind sight is that the blind person is still sensitive to the information emanating from the source, even if it is not connected up with visual sensation.

The hypothesis in the case of the Prospagnosic who passes the guilty knowledge test is that she is relying on an aspect of face recognition unconnected with the visual system. Bauer suggests that there is more than one route to recognition of familiar people. There is an intellectual one, in which visual and semantic information is processed, and a noncognitive, emotional, one. The
former, ventral route runs from the visual cortex to the temporal lobes via the interior longitudinal fasciculus. However recognition of familiar people usually exploits a parallel process which runs by the dorsal route from the visual cortex to the limbic system via the inferior parietal lobe. The limbic system is the organic basis for many emotional states and responses. Bauer’s hypothesis, developed by Ellis and Young, is that when we recognise familiar faces, perhaps even familiar situations or locations, we do so via both routes.

The prosopagnosic who is unable to recognise familiar faces yet still knows who the people are, may be relying in this emotional response even though the intellectual route is totally or partially destroyed. So the sensation of familiarity remains even though the sufferer cannot detect a basis for it.

In Capgras syndrome the situation is reversed. The patient has all the information her cognitive system needs for recognition but she feels no sensation of recognition because the route to the limbic system is damaged. This may have something to so with the paranoiac aspect of the disorder in which people feel that they are being preyed upon by imitators and strangers whose impersonations make them indistinguishable from familiar friends and relatives. The absence of the sensation of familiarity is the origin of this unease. The fact that Capgras sufferers only feel this way about familiar people and have a similar sense of derealisation with familiar scenes and situations tends to reinforce this hypothesis.

I do not wish to suggest that the explanation of autism has the same organic base as Capgras syndrome. Rather I wish to emphasise how closely the cognitive, emotional and affective aspects of recognition are tied to one another. The bizarre phenomenon of Capgras syndrome shows how severe is the damage to the normal process of interpersonal recognition when this link breaks down completely in adulthood, after the familiar patterns of recognition have been ingrained over many years. The cases of prosopagnosia and Capgras syndrome lend a posteriori support to the intuitions on which Wittgenstein’s claim relied. So, while it is not a priori true
that recognising another's mental states requires an element of emotional response, it seems to be a contingent feature of human psychology that emotional responses normally play a role in some acts of recognition crucial to the construction of contexts for social interaction.

5.2.1. The Role of Emotion in Context Construction

The autistic, although able to process the visual and other information, never seems, even in infancy, to connect it up with sensations of affective response to other people. When the normal baby plays its imitation games, responds to its mother's smiles and seeks out her glance she is enjoying an affective interchange. Why would the infant bother to imitate, catch its mother's eye and exchange smiles in the absence of the affective sensation of interpersonal engagement which accompanies these activities? Someone who lacks these basic connections will be unable to construct a context for even the most basic acts of intentional communication. The context for these early exchanges depends on the ability to respond to the mother's emotion as conveyed by her expressions and gestures. My suggestion will be that while the normal child develops its mindreading skills by employing a battery of capacities organised in concert with its conscious or unconscious affective responses, the autistic child lacks this affective component of social recognition.

5.3. Emotion and Somatic Markers

The thesis I want to examine is that the basis for defective intentional understanding in autism is more likely to be an inability to integrate affective response with memory, reasoning and attention, than a purely cognitive malfunction. I suggest that the normal integration of those capacities is not via the...
functioning of a module devoted to the production of M-representations. Furthermore, I want to link this hypothesis to the empirical evidence about the organic bases for social recognition and response.

One cannot test my hypothesis by damaging the areas of the brain connected with emotion in an attempt to produce autistic symptoms. One has to rely on other neurological evidence about brain function and the best theory explaining the neurological evidence about the interconnection between emotion and cognition in social contexts.

As a preliminary step in my argument consider what we know about the organic basis of affect. It has been known for a long time that damage to particular areas of the brain has the effect of reducing or destroying completely a patient's affective responses. I shall briefly develop a model, derived from Damasio, of the brain's function which would explain this connection between affect and specific areas of the brain.

The first and most evolutionarily basic of the brain's functions is the homeostatic regulation of the body. That is keeping body temperature, heartbeat, oxygen supply, blood sugar and the myriad of chemical interactions on which the functioning of the body depends, within the fairly narrow range which ensure survival. To do so the brain constantly monitors body state. This monitoring is not, however, a matter of the receiving of neural signals from distant parts, their integration in a control centre, the formulation of a response, and the transmission of neural signals which produce an updated response. The sections of the brain which represent and control the states of various body systems for homeostatic control do so without being linked in a central processing unit. Another aspect of the homeostatic regulation is that it is not produced solely via neural signals. For instance, certain areas of the brain are responsible for the release of hormones which are carried by the bloodstream to the appropriate site. The results of

4 (Bowers 1993, Damasio 1994) Chapter 4 contains an introduction to the neurological literature on affect.
hormonal activity at that site may be signalled to the brain neurally or by information carried by the bloodstream back to the brain. The brain also produces neurotransmitters, chemicals which increase or decrease the level of excitation in specific neuronal circuits.

This type of homeostatic regulation depends on representation by the brain, as levels of activation in dedicated neuronal circuitry, of the information constantly provided via different channels. Responses to perceptual input exploit this type of regulation system: for example the sight of a predator produces a flood of adrenalin, perspiration and raised heartbeat. These are the bodily responses we experience as fear. As William James put it:

If we fancy some strong emotion and then try to abstract from our consciousness of it all the feelings of its bodily symptoms, we find we have nothing left behind, no "mind stuff" out of which the emotion can be constituted and that a cold and neutral state of intellectual perception is all that remains... What kind of an emotion of fear would be left if the feeling of quickened heartbeats nor of shallow breathing neither of trembling lips nor of weakened limbs, neither of gooseflesh nor of visceral stirrings, were present it is impossible to imagine. 5

Obviously we do not experience all the body states involved in homeostatic regulation. The alteration of chemical balance in our bodies and bloodstreams, the operation of our nervous system and the release or suppression of neurotransmitters are not normally present to consciousness. However in many cases we are aware in consciousness of altered body states. Fear, sexual arousal, sadness, enhanced attention are examples. These are primary emotions. 6

5(James 1950) quoted by Damasio at p129.
6Damasio reserves the term, "primary emotion" for the changes in neural and body state. He calls the conscious awareness that accompanies such changes "feelings." As long as the distinction between those changes occurring below the level of consciousness and those which give rise to conscious sensation, is recognised, terminology does not matter. I have used primary emotion and secondary emotion to describe responses which may or may not be conscious.
In many species, including humans, some body states are innately connected to representations of specific environmental features. So size (as in large animals) type of motion or certain sounds (as in growling or the comforting sounds made by mothers to babies) are automatically connected up with dedicated neural circuitry which controls an autonomic response through either the endocrine or nervous system. The relevant neural circuitry is located in the prefrontal area of the brain (more specifically the amygdala and the anterior cingulate regions). These are areas of the limbic system: that part of the brain concerned with homeostatic regulation whose structure is essentially shared with other organisms. Primary emotions exploit the homeostatic regulation system to produce an advantageous response. Specific types of perceptual stimuli alter body state so that the appropriate response, mating behaviour, flight etc. occurs. The knowledge on which these responses depend is held as a series of what Damasio calls dispositional representations: sets of neuronal circuits with an innate propensity to become activated as the result of the representation of specific bodily and perceptual states.

All knowledge is the result of the dispositions of certain representations (patterns of excitation in neural circuits) to become active as the result of excitations of other circuits. These patterns of excitation are the result of the transmission of information via neural signals, neurotransmitters and modulators, and hormones carried by the bloodstream. The circuitry which supports primary emotions is now, as a result of evolution, genetically hard-wired.

Secondary emotions are those which are experienced as a result of representations which are not genetically determined, but acquired, and which exploit the same neural circuitry as primary emotions. For instance the feeling of nervous anticipation or the sadness evoked by imagining certain situations, are inspired by the activation of dispositional representations in the early sensory cortices. These circuits are disposed to reconstruct, more or less accurately,
perceptual representations of those situations. The presence of these representations is signalled to the amygdala and anterior cingulate and, as a result, dispositional representations in these regions are activated. Their functions are: signalling to the body via the autonomic nervous system so that the viscera are placed in the state most commonly associated with the recalled situation; dispatching signals to the motor system to place the body in appropriate attitude; activating the endocrine and peptide systems which transmit signals via the bloodstream; and, finally, activating neurotransmitter nuclei in the basal forebrain which then release their varied chemical messages in, for example, the basal ganglia and cerebral cortex. The resultant body state is instantly signalled and monitored by the same complex system.

There are several points to note. The first is that secondary emotions do not depend on perception; they depend on the activation of "as if" representations sustained in the somatosensory cortices. The second is that secondary emotions exploit the same systems as primary emotions. The body does not have one system of emotions connected with bodily function in actual situations and another connected with more abstract representations of "as if" emotions. The secondary emotions depend on the same system of homeostatic regulation as the primary. The third is that these bodily changes are unconscious and automatic even though our experience of them is, in many cases, registered as a change in conscious emotional state.

The fourth and most important is that this view of the relation of brain and body contradicts the conventional wisdom about the connection between reasoning and emotion or intellectual and bodily function that the brain is like Descartes' pilot within a vessel: receiving messages from all over the body, planning a response and sending back messages via the nervous system. However there is no site in the brain at which all this information converges. There is no central processor which receives input from the periphery, evaluates it, selects an action and sends out
appropriate instructions. Another way to put this point might be to say that there is no site in the brain which contains a map of the organism, constantly updates that map and "decides" what the organism should do next as a result. Rather, areas of the brain vary in state according to variations in states of the organism. These areas and the bodily states they monitor are connected by an elaborate system of feed forward and feedback loops exploiting neural, hormonal and neurotransmission signals, but there is no central "exchange" through which these signals are all routed. This is Damasio's description of the process.

It is true that there are a few brain regions where signals from many different early sensory regions can converge. A few of those convergence regions actually receive a wide variety of polymodal signals, for instance the entorhinal and perhinal cortices. But the kind of integration those regions can produce using such signals is unlikely to be the one that forms the base for the integrated mind. For one thing damage to those higher order convergence regions, even when it occurs in both hemispheres, does not preclude "mind" integration at all, although it causes other detectable neuropsychological consequences such as learning impairments. It is perhaps more fruitful to think that our strong sense of mind integration is created from the concerted action of large scale systems by synchronizing sets of neural activity in separate brain regions, in effect a trick of timing....

The fundamental problem created by time binding has to do with the requirement for maintaining focused activity at different sites for as long as necessary for meaningful combinations to be made and for reasoning and decision making to take place. In other words, time binding requires powerful and effective mechanisms of attention and working memory....

Each sensory system appears equipped to provide its own local attention and working memory; human studies as well as animal experiments suggest that the prefrontal cortices and some limbic system structures (the anterior cingulate) are essential.\footnote{(Damasio 1994 p94)}
It should be apparent that if this model is correct another common belief about the brain's operation is incorrect. That idea recue has it that the intellectual aspects of the brain's function are localised in the frontal cortices while the neural substrate of the emotions is confined to the limbic system, that part of the brain devoted to homeostatic regulation, whose structure we share with other animals. Damasio proposes, instead, that reasoning cannot take place without involving those parts of the brain devoted to bodily regulation. This relation between emotion and reasoning is in fact integral to learning and survival and to the acquisition of social and cultural attitudes. This is the "somatic marker" hypothesis, of which I present a truncated (and freely adapted) exposition below.

Before evaluating this hypothesis, which is largely an interpretation of evidence about the neurological localisation of emotional response and the effect on social cognition of damage to these locations, an important caveat should be registered. Neurological localisation does not prove modularity although it may help in determining which dedicated neural circuitry supports a modular function. Modularity is a thesis about informational encapsulation. We know for example that much of the information processed by the visual system is unaffected by information which reaches the organism by other channels. It is this which makes vision modular, not the fact that it takes place in certain areas of the brain. Conversely, showing that more than one area of the brain is involved in a certain activity does not show that the information required for that activity is not encapsulated by a single module. However the account of somatic marking I give is consistent with the hypothesis that context construction need not depend on the functioning of a single module which integrates the output of other, modularised, perceptual and affective systems via the representation and application of theoretical principles.

It makes evolutionary sense, firstly, that perceptual representations are of information salient to the survival of the organism and, secondly, that those
representations can have appropriate effects on the homeostatic regulation system. Thus the perception of a predator-like shape has almost instantaneous effects on the body state of the predated species. Certain perceptual representations are *somatically marked*, linked via the operation of dispositional representations, to the bodily regulation system. The connection between perception and survival mechanisms in these primary cases is innate: the dispositional representations are the result of evolution.

However, in other species, the connection between perception and advantageous behaviour is not so automatic. These organisms have the ability to learn from experience and adapt their behaviour accordingly. The point to note is that the brain continuously represents body state, and thus primary and secondary emotions, conscious or not, during the learning period. Those representations which are strongly somatically marked, associated by dispositional representations with favourable or unfavourable primary or secondary emotions, are those most easily retrieved and held in focus by working memory. Learning thus depends, not merely on the cognitive and intellectual aspects of memory and attention, but on the somatic marking of particular representations, reinforced by the process of development.

Planning or deliberating is a matter of responding appropriately to a given contingency. That contingency can be represented as the result of perception, memory or imagination. Possible responses are also represented and an alternative is chosen and acted upon. However, to echo a point made in the second chapter, not all possible responses are represented: deliberation is a matter of narrowing possibilities, often to the stage where the link between perception and action looks and feels automatic. This immediacy is a product of somatic marking: of the possible responses some are marked out as more salient than others via their association with somatically marked memory representations. The very fact that
these memories are themselves somatically marked in certain ways ensures that they are the ones held in working memory during evaluation.

Organisms act on the basis of skills, including the ability to reason explicitly, acquired over the course of particular lives which are played out against cultural and social backgrounds. Which representations are somatically marked for individuals will thus vary according to primary or secondary emotions prevailing in the context in which they were acquired.

The result of the somatic marker hypothesis is that we should expect that secondary emotion is crucial in the adjustment of action to perception, whether or not this adjustment is the product of conscious deliberation. Another consequence of the hypothesis is that it is reasonable to suppose that those representations somatically marked among the higher primates would include those conducive to negotiation of the social environment. Many primate secondary emotions should thus be linked to typical patterns of social interaction.

So far I have not given any evidence for the somatic marker hypothesis, merely summarised it. Before I give some of that evidence and relate the overall account to the case of autism one more point needs to be made. According to Damasio the dispositional representations which control secondary emotion active in personal planning and social interaction depend on the neural circuitry of the prefrontal cortices, particularly the ventromedial sector. Circuitry in these sectors seems to be paired with circuitry which underlies perception in all modalities. Recall the case of Capgras syndrome and prosopagnosia. The explanation I endorsed was that they were mirror images of each other, breakdowns of the connection between affective and cognitive aspects of face and personal identity recognition. If Damasio is correct, then we might expect further neurological investigation to refine Bauer's hypothesis that face recognition depends on representations of visual, personal identity and affective information. That affective information is best conceived of as an unconscious secondary emotion.
which grounds the feeling of familiarity in prosopagnosia and whose absence produces the paranoia characteristic of Capgras' syndrome. Another way of refining the account I gave suggests itself. The discussion on which I relied models face recognition as the convergence of information on a personal identity node, as if other subsystems fed information into this central processor which is then connected to a name generator. Damasio's model suggests, rather, that face recognition requires a series of representations including secondary emotion, to be entertained simultaneously in parallel whenever a face is recognised.

5.3.1. Evidence of Somatic Marking

Damasio's hypothesis draws much of its plausibility from evidence about the result of damage to the prefrontal cortices. Monkeys are not humans, or even chimpanzees, but they do live in societies which depend on the coordination of behaviour via the interpretation of movement and expression. Myers reported that monkeys with bilateral prefrontal ablations, damage to those areas of the brain which Damasio suggests are crucial to the integration of affect in social understanding, do not maintain normal social relations within the group.\(^8\) These brain-damaged monkeys no longer groom themselves and others, have diminished facial expression, impaired maternal behaviour and sexual indifference. Their movements and physical abilities are unimpaired but their social relationships within the troupe are destroyed as a result of their brain damage. It is not just that they no longer relate to other monkeys in the troupe in the normal way, but the other monkeys cease to include them in the normal social behaviour of the tribe such as grooming rituals and displays of dominance. Damasio concludes his survey of this and other evidence on the connection between brain damage and social cognition in primates as follows:

\(^8\)(Myers 1975)
It is fair to assume that monkeys with prefrontal damage can no longer follow the complex social conventions of the organisation of a monkey troop. It is likely that they fail in terms of "social cognition" and in terms of "social behaviour" and that other animals respond in kind. Remarkably monkeys with damage in the motor cortex, but not in the prefrontal cortex, have no such difficulties.

Damasio goes on to conclude:

In spite of the marked neurological differences between monkey and chimpanzee, and between chimpanzee and human, there is a shared essence to the defect caused by prefrontal damage: Personal and social behaviour is severely compromised.\(^9\)

Damasio's other examples are of people with damage to the prefrontal cortices, but in whom working memory, learning, attention, language and reasoning remain intact. In all these cases social decision making and responses are impaired, accompanied by a diminution of affect. This suggests to Damasio that there is more to deliberation in social contexts than the purely cognitive evaluation of alternatives. What has been lost is the ability to mark somatically some alternatives as preferable to others. More noticeably, damage to the prefrontal cortices results in personality change as people no longer take an interest in friends and relations or the trajectory of their own life.

In each case of damage to the prefrontal cortices observed by Damasio the patients have noticeably flattened affect and defective social decision making. By defective social decision making Damasio means an inability to choose an advantageous course of action (the action acknowledged by reason as most

\(^9\)(Damasio 1994 p75)
advantageous) or to continue to live in a way consistent with previous social identifications and affiliations.

Damasio also relies on the phenomenon of anosognosia in support of his hypothesis about the connection between emotion and direct bodily monitoring. Anosognosia, normally coming about as the result of a stroke which paralyses one side of the body, is the inability of patients to recognise their own illness via the homeostatic regulation system, precisely because that system is damaged and thus does not represent the state of their paralysed limbs. Left sided paralysis, caused by a specific pattern of brain damage is accompanied by anosognosia, right sided paralysis is not.

Anosognosic patients simply do not feel impaired. When asked how they are, they reply "well" even though their limbs are paralysed. This is not just denial, because they can acknowledge, intellectually, that they are indeed paralysed when it is pointed out to them by a doctor. The same patients are unworried by their condition; as Damasio says "emotion and feeling are nowhere to be found in anosognisic patients, and perhaps this is the only felicitous aspect of their otherwise tragic condition. Perhaps it is no surprise that these patients' planning for the future, their personal and social decision making, is profoundly impaired. Paralysis is the least of their problems." 10 In contrast, patients with right side paralysis are acutely aware of their paralysis, with consequent depression and suicidal worry about their future and the effect on their families. Anosognosia seems a perfect example of the connection between monitoring of the body state, primary and secondary emotion, and their interrelationship in the reasoning required in social contexts.

Or consider some of the failures of rationality often remarked by Kahneman and Tversky. People will choose to have an operation if told there is a ninety percent success rate but not to have the operation if told there is a ten percent death

10(Damasio 1994 p64)
rate. This is of course unsurprising if the alternatives are somatically marked differently despite their identical probability. Consideration of the ninety per cent chance is accompanied by representations of health and happiness. Mention of a ten per cent chance of death produces representations of death and desolation correlated with anxiety and fear. As a result the homeostatic regulation system produces analogues of these states which are instantaneously monitored and feed back into the decision process. In considering the ten per cent possibility of death the patient feels surrogate experiences, secondary emotions, of fear and bereavement. Thus she will choose not to have the operation. None of these processes or the secondary emotions need be conscious. Perhaps all the patient is conscious of is the preference not to have the operation after what seems to her a clinical consideration of the probability.

Damasio's contention seems to be borne out by skin conductance measurement carried out on patients with damage to prefrontal cortices.¹¹ As I mentioned in the discussion of Capgras syndrome, skin conductance response is a way of detecting changes in homeostatic regulation which may not be experienced consciously. Damasio's experimenters showed normal patients, patients with prefrontal damage but otherwise intact mental faculties, and other patients with brain damage but intact prefrontal cortices hundreds of slides of social situations including disturbing, frightening and emotionally upsetting scenes. The normal patients and those brain damaged patients with intact prefrontal cortices had increased SCRs for those disturbing pictures, but the patients with prefrontal damage showed no variation. The testimony of some of these patients confirmed that, although they could recognise these scenes as those which typically compelled a certain response such as horror or pity, since their brain damage, they no longer felt such responses.

¹¹(Damasio 1990)
This finding comports with the classic cases of patients with prefrontal damage: they are unable to link affective considerations to consideration of their own social situation and their patterns of affective response to others, especially familiars, are disrupted.

5.4. Context Construction and Somatic Marking: The Autistic Deficits

What are the consequences of Damasio's findings for the ToMM explanation of normal and abnormal child development, particularly autism? Autism is a developmental disability resulting primarily in the failure of the ability to understand the meaning of others' words and actions. I proposed that we should understand the autistic person as one who is unable, even from the earliest stages to construct a context, in Sperber and Wilson's technical sense of the term, for even the earliest and most basic communicative acts between a child and its mother.

Rather than explaining the autistic child's deficits in terms of the failure of a module which generates theoretical interpretations of behavioural evidence, I propose to revive Kanner's original thesis about autism: that the difference between autistic and normal children is primarily an affective deficit. Of course this is not necessarily inconsistent with the ToMM model. It could be the case that autistic children have no affective drives to engage with others and thus no impulse to develop a theory of other minds. In Baron Cohen's terms, emotional engagement could be the key which unlocks successive precursive stages of ToMM. Because autistic children are not interested in other people they have no impetus to develop a sequence of more sophisticated theories about the reasons for their behaviour.

12Someone who has consistently argued, against the ToMM model, for this way of understanding autism is Hobson. His arguments advance considerations similar to those advanced by Wittgenstein and do not consider possible organic bases of the deficit. (Hobson 1991)
This theory of the link between affect and context construction can also be applied to accounts which replace the ToMM with another module such as a simulator. The idea would be that as a result of its need for affective engagement the child begins to interpret others' behaviour by simulating their mental states in progressively more sophisticated ways. A version of this theory, without any commitment to the simulation model, but which emphasises the interconnection of affect and imagination in the child’s fantasy life can be found in Mayes et al.\textsuperscript{13}

However there is no single mechanism, whether a simulator or ToMM, responsible for intentional understanding, which is "switched on" by the innate drive for affective contact. Rather, intentional understanding requires the representation of a context for the understanding of words and gestures. The elements of that context are represented in parallel. Also represented as part of the contextual array, are the primary and secondary emotions associated with contextual representations. So, for example, if understanding another's words requires the recognition that they are spoken in anger by the father, then the father’s face is represented as is his expression, posture and tone of voice along with the primary and secondary emotions like fear.

The mere fact that information is represented in parallel does not automatically defeat claims of theoreticity or modularity of processing. However in this case it does seem to show that, in the normal case, the information relevant to social cognition is not theoretically integrated via the application of some explicitly represented principles. Whether or not social cognition is modularised depends on the whether or not the affective and cognitive information is encapsulated together. My interpretation of Damasio's account is that it is not. Instead there are many channels of information each of which is encapsulated and whose output is integrated in the way Damasio suggests, by being represented in parallel. The importance of the somatic marker hypothesis is that it explains why

\textsuperscript{13}(Cohen 1993 p461)
some information is represented more strongly (held in working memory and
attention) than others without the need to invoke a set of explanatory principles to
govern the relative weighting of representations.

The case of Capgras' syndrome showed just how closely normal acts of
recognition are bound up with affect. Consider my recognition that another person
is angry. On Damasio's account this is the result, not only of computing visual
information about posture and expression but of the parallel representation of the
appropriate secondary emotion. That anger may be an important element of a
conversational context, whose elements will include the semantic content of any
words actually spoken and representations of salient background information, the
possible causes of the anger etc. As a result of the simultaneous representation of
all these contextual elements an agent is able to decide what is communicated by
the words and to formulate a response. That response may be as automatic as a
flinch or as calculated as a diversion of the conversation to more neutral topics.
Whatever the response, the representations which accompany deliberation will
have been somatically marked.

According to Damasio's account, all responses which require the
construction of a context involve integration of primary or secondary emotions
with the representation of other contextual elements. If this is correct, then an
explanation of the automaticity of responses produced by acculturation and social
experience is suggested. People who have acquired knowledge in particular
situations have done so by acquiring somatically-marked representations which
are then stored in memory. The later availability of those representations to
working memory and attention depends on the strength and nature of conscious
or unconscious primary or secondary emotions.

My surmise, in two parts, is, firstly that autism is a failure, from the very
outset of the infant's life, to somatically mark the information necessary to
construct a context. Secondly constructing a context is a matter of synchronising a
variety of representations of contextual elements, represented by different modules, rather than channelling them through a hypothetico-deductive inference engine which generates M- representations. The responses which flow from that synchronic representation are the activities which we classify in propositional attitude vocabulary. So we can say, on the basis of what they say and do, that an agent is frightened of another person’s anger or believes that her friend is unhappy or embarrassed. However, in so doing we do not need to infer that the agent has made, and internally represented, some hypotheses about the nature of the representations in the brain of her friend based on a theory of the social world she has acquired from experience.

Of course we do learn from experience of the social world, by representing the information we acquire via a variety of perceptual and other modules (some of which will be discussed in the next chapter). These representations are somatically marked and accrete in memory as dispositional representations available to be linked with new perceptual information, other memories, or the circuitry which sustains imagination and deliberation. The nature of these linkages is not entirely known but it seems that the assembling of representations in response to new information is governed by somatic markers.

This account also has the potential to explain some other puzzling aspects of autism. The stereotypical, repetitive behaviour and the fascination with concrete details of the physical environment, for example, can be seen as the result of idiosyncratic somatic marking. Where the normal child somatically marks aspects of the social environment, and those behaviours which immerse her further in it, the autistic child somatically marks aspects of the physical environment.

Somatic markers may also be relevant in explaining what some have characterised as an executive function deficit. This has been recently proposed as an alternative to ToMM or Simulation accounts as a unifying explanation of the
variety of autistic behaviours. Without endorsing the account, one can point out that Frith and Happe are right to note that a feature of autism is the repetitive focus on singular concrete instances. It could be, as they suggest, that this focus is a result of an inability to organise information schematically by generalising and abstracting from particular instances to form a coherent projectible account of the domain in question. This account has a number of attractive features: its compatibility with the emphasis by ToMM on the integration of information needed for context construction; its relevance to the repetitive and stereotypical play, and the way autistic abilities and interests are not distributed across different domains.

If certain activities or objects are idiosyncratically and strongly marked to the exclusion of similar objects and activities, then one dimension of similarity which could ground abstraction is lost. Alternatively it could be the case that, due to the abnormal function of somatic markers, the autistic child, by concentrating on particulars, never develops the impulse to form an integrated account of her circumstances. This may be especially true in the social realm. The autistic child seems to have no motive to try and put together a context for the interpretation of behaviour.

These remarks are not intended to reduce all the competing accounts of autism and normal development, ToMM, Simulation, Executive Function or the Empathetic account to one: somatic markers. Rather they emphasise a feature common to all these accounts: they focus on capacities sustained by representations which are somatically marked.

On my account a univocal account of normal and deviant development in terms of a single mechanism or module is not required. What is needed in the case of the child’s understanding of other minds is to explain how the child acquires the dispositions we come to classify using intentional vocabulary. Some of the abilities

\[14(\text{Frith 1994})\]
the child needs to construct a context for interpretation of other minds seem to be innate: the connection of primary emotion with certain facial expressions and auditory stimuli. Others depend on the development of innate dispositions to fantasise and play imaginatively. Others may depend on executive function, others on language learning. In some cases we may even rely on M-representation, but there is no reason to suppose that this variety of abilities is routed through a single module devoted to other minds. It seems more likely that, since we are creatures whose survival depends on our ability to cooperate and communicate, we employ all the capacities with which natural selection and learning endow us to that end. We do so by the parallel representation of information from a variety of sources. Crucial to the selection, display and memorising of that information is the correct functioning of somatic markers: hence my hypothesis that one of the primary deficits in autism is in the somatic marking of information relevant to context construction.
5.5. Philosophical Connections

In the first chapter I followed Churchland, Pettit, Dennett and others who characterise the practice of intentional psychology as a species of pattern recognition. (See 1.4.) Thus, someone who understands the crucial intentional concepts of belief and desire can apply them to subsume instances of behaviour under the general pattern. For someone who understands intentional concepts, human behaviour is laden with meaning because it conforms to patterns which can only be recognised, as Dennett puts it, from the intentional stance. Recall the experiment Baron-Cohen devised which presented children with a series of scenarios which can be interpreted as a sequence of causally related events in the natural world, like someone pushing a rock off a cliff, interpreted behaviouristically, or as a series of actions and responses interpretable in intentional terms. It is this latter ability, the ability to "see" patterns of behaviour as intentional action, which is practised automatically and without conscious effort, which theorists of developmental psychology need to explain.

A parallel might be the acquisition of language, which clearly depends, initially, on the ability to hear speech, not as a random series of sounds but as organised according to some underlying principles. Just as the generative grammarian postulates an innate ability to interpret human speech via the application of fundamental grammatical principles represented in the human brain, ToMM explains the acquisition of intentional concepts in terms of some innate capacities for the interpretation of human behaviour. In both cases, activity is categorised as falling into a pattern via the application of domain specific concepts, grammatical or psychological.
Exactly how the mind accomplishes this recognition is an empirical question. Perhaps, as the ToMM theorist suggests, the ability to "see" behaviour as intentional, as the activity of minded agents, depends on the acquisition of a theory of the domain in question. I have argued the evidence about the acquisition of intentional concepts can be explained without postulating a modular capacity of the brain which explicitly represents rules for the application of psychological concepts.

My suggestion in the autistic case was that autistic children fail to develop mindreading abilities because the relevant representations are not somatically marked. Another way to put this is to say that certain patterns of behaviour never become salient to them because of their lack of certain crucial emotional responses. Note that this claim is compatible with the truth of either ToMM or the competitor theory I advanced. It could be the case, for example, that in virtue of a lack of emotional response to others and the deviant somatic marking of representations of the physical environment, the ToMM module never gets "switched on".

In either case, the acquired ability to apply intentional concepts, to recognise behaviour as intelligible activity, will depend on the emotional responses to the actions of others.

The idea that the ability to apply certain concepts depends on a particular emotional or noncognitive response is familiar in other domains. Qualitative and evaluative or moral concepts are often analysed this way. Take, for example, the concept of embarrassment. There are at least two ways in which it can be recognised. The first way, the functional way, is to treat embarrassment as a state we recognise by the presence of certain behavioural types, blushing, agitation or displacement activity. It does not matter if we then identify the mental state of embarrassment with the causes of these behaviours or, in the Rylean way, with the behaviours themselves. The point to note is that one can judge that someone else is embarrassed purely by observing their behaviour. Someone who had never been
embarrassed, or felt sympathy for someone else's embarrassment would, on this account, be able to detect embarrassment via the operation of a purely behavioural or functional theory.

However, we also might recognise embarrassment empathetically, by imaginatively projecting ourself into the situation of another. If the imaginative identification produces a flushing of the skin and a visceral unease we might then proceed to the judgement that she is embarrassed. Of course in most cases the simulation, and perhaps even the response, might be unconscious and automatic. Nonetheless the judgement would depend on the ability to experience embarrassment from the first person perspective.\(^{15}\)

Still another possibility is that the perception of embarrassing scenes or behaviours causes an analog of embarrassment in the perceiver, just as as the sight of a mother's smile produces a pleasant sensation in a baby.

In both these cases the ability to recognise embarrassment depends on the ability to feel embarrassed or to experience the typical emotional response to another's situation.

This is not to deny the possibility that embarrassment can be recognised in the functional way; it is just to claim that in the normal case our response to another's embarrassment is mediated by an analogy of the emotion in ourselves.

To make the case more vivid consider the case of colour vision. Suppose that we identify colour with one of the common candidate properties, a spectral reflectance pattern, or the activation of neuronal circuitry in the visual cortex. In this case, to use the vocabulary of the functionalist, the realiser of the functional role of colour will be the physical property which contingently plays a certain functional role: namely causing judgements of the form "That is red."\(^{16}\) However it seems logically and nomologically possible that our ability to recognise colours

\(^{15}\)Nicholas Humphrey has suggested that conscious experience confers an evolutionary advantage to primates who need to function in complex societies.

\(^{16}\) This way of putting the point is borrowed from Pettit and Jackson. (Jackson 1988)
may depend on a type of experience, which can be had even in the absence of the preferred physical realiser, namely the sensation of seeing red, orange or vermilion.

In the normal case colour experience and the presence of colour are present together. It may even be the case that the presence of colours (spectral reflectance properties or firing of neuronal circuitry) causes colour experience. However, such causal relations may break down, as in cases of colour blindness or the loss of colour sensation after a stroke. Such people may still be able to detect the presence of colour but not experience the sensation which normally goes along with seeing, for example, sunsets and fire engines. Colour blind people, for instance, are often good at coordinating their colour judgements with those of normally sighted people: they are responding to the same physical property as the normally sighted but their responses are not mediated by the same experience. Another example is the phenomenon of blind sight in which people who are clinically blind, and who report no visual experience, nonetheless do far better than chance at detecting the presence of objects in their visual field.

Of course in bringing their judgements into line with the rest of the community, the colour blind rely on the judgements of normally sighted perceivers, made on the basis of colour experience. The same would be true of anyone who learnt to apply a colour concept purely on the basis of functional or physical properties without being able to rely on the normal sensational experience.

Moral or evaluative concepts can be analysed in a similar way. The analysis starts from a familiar fact about moral properties: judgements about whether their predicates apply supervene on the nonmoral facts which attract the judgement. This supervenience principle holds that one cannot make differing moral judgements about two situations whose nonmoral properties are identical. So that one cannot, for example, regard one instance of jumping into shark-infested waters
as courageous and another as reckless unless they differ in some relevant respect. If they are identical in all respects but are judged to be morally different then either the judger does not understand the nature of moral theory, or she has changed her theory in the meantime, her theory is inadequate or ambiguous, or the judgements are not really statements about the moral properties of the situation but about the variable psychological states of the judger. This supervenience of the moral on the non moral is an *a priori* truth about morality which the theorist needs to explain.

The first step seems to be to notice that it is possible to distinguish the moral from the nonmoral features of a situation. Otherwise one could not convict another of inconsistency in judgements. Inconsistency arises when someone makes different judgements about situations which are identical in all relevant respects. Another way to make the point is to note that different cultures or different people have different moral attitudes to the same types of situation. What Ludwig Wittgenstein thought was an unexceptionable means of primary school education, Pavlovian rote learning reinforced by severe corporal punishment, is considered child abuse by present day educators. Note, however, that they would agree about the non moral features of the situation, the raining of blows on the head of a child who gets her sums wrong. Wittgenstein, however, considered this a morally neutral aspect of the education process, the modern parent considers it an instance of sadism.

The difference appears to be in their responses to the situation. There are a variety of ways to explain the relation between moral judgements, responses and the nonmoral features of a situation. (This section is not an essay in moral philosophy or an argument for a particular position, so the possibilities I sketch are not meant to serve as arguments for a particular moral theory any more than the section on Aristotle in chapter two was intended as an argument for his version of the acquisition of ethical knowledge.)
One possibility is that particular situations compel certain responses in humans, such as fear, sympathy, anxiety or pleasure, and that our moral judgements are culturally mediated extensions and articulations of these basic responses. This implies that the normal ability to make moral judgements depends on the presence of these responses, just as the ability to apply colour concepts requires the presence of colour experience. If this is correct then we must recognise the possibility of an analogue of colour blindness for moral concepts, that is to say; that there may be people who lack the relevant response but who can, nevertheless, make the same judgements as the rest of the community by carefully observing which nonmoral situations attract typical moral judgements and learning, in an alienated way, to bring their own judgements into line with the rest of the community.

There are at least two ways to analyse this apparent fact about moral judgements, the first is to say that the person who judges in the alienated way is not making moral judgements at all, but sociological judgements about the things people typically do in certain circumstances. Such a person is not psychologically equipped to perceive situations as good or bad because although they can correctly predict what others will say or do in a particular situation they are not recognising the situation as morally significant via the normal response.

The second is to say that what the possibility of alienated judgements shows is that moral judgements are not really cognitive, that sentences like "courage is a good thing" do not state truths or falsehoods about particular actions or situations: they merely express feelings typically aroused by those actions or situations.

Whether one analyses moral judgements in an expressivist or cognitive fashion does not really matter to this discussion. The salient point is that both positions start from the premise that the normal use of moral concepts somehow involves a response evoked by the situations in which the concepts are applied.
Those who lack this typical response may learn to apply the concept in an alienated way, but they do not understand the concept in the same way as someone who learns and applies it on the basis of the typical response. The alienated version of agoraphobia for instance, in which people understand that open spaces make certain people afraid may not give much insight into the nature of the agoraphobic experience. Similarly, it seems unlikely that we get much insight into the life of a schizophrenic or seriously depressed person by an inventory of their symptoms and their typical causes. The alienated perspective may be vital for diagnostic purposes, but there is another, experiential, way of understanding the condition which is not captured from the alienated perspective.

So far I have only argued that there appear to be two ways of understanding and applying concepts, the experiential and the alienated. If this is correct then the possibility that the acquisition of psychological concepts may depend on certain innate experiential responses which ground dispositions to engage intersubjectively is not surprising. Intentional concepts will be just a subset of the wider set of concepts whose normal acquisition depends on the ability to experience a typical response. This is not to deny that psychological concepts can be learnt and applied in the alienated way, rather it points out that in the normal case our ability to learn and apply psychological concepts goes hand in hand with feelings of affective engagement with others.

Wittgenstein, of course, argued that the experiential perspective was necessary in order to understand and apply the concept. So that, for example, one could not judge that someone else was embarrassed or happy unless that mental state produced an analog experience or response. So, for Wittgenstein, someone who made their judgements about others in the alienated way was not really

---

17 Perhaps this is an overinterpretation of Wittgenstein’s remarks on the lack of inference involved in the phenomenology of perception. Nonetheless I think that Wittgenstein would have agreed with William James remark and, furthermore, that the alienated notion of understanding was not part of our everyday concept of the grasp of a concept with an emotional content. Because the alienated way of understanding does not depend on agreement in a form of life.
applying intentional concepts at all but engaging in what William James called the "cold and neutral state of intellectual perception." This, I think is the point of his remark that responding to others' mental states is not a matter of making an inference. He thought of inference in this context as analogous to scientific theorising about unobserved causes.

However, it is not necessary to the understanding of intentional concepts that one possess a particular response. I argued that affective responses make the patterns interpreted and explained by intentional concepts more salient, especially in early development. It is a contingent feature of human psychology, rather than a necessary feature of intentional concepts, that our recognition of the situations to which those concepts apply is enhanced by the presence of innate emotional responses. To use Damasio's language, certain representations of others' behaviour, especially behaviour expressive of emotional states, are somatically marked. These somatic markers enable us to see those behaviours as expressive of mental states rather than as non-intentional bodily movements.

The application of intentional concepts is an instance of a general phenomenon discussed in chapter two. When confronted by a domain of activity the human mind does not generate alternatives by ranging over all the possible explanations. Rather we apply domain-specific concepts which limit the possible explanations of the phenomenon. The ability to apply these concepts may be innate or it may be acquired. In either case, the phenomenology is of an immediate recognition, or perceptual gestalt, that the situation is one to which a particular concept applies. The expert radiologist who examines an X-ray or an ultrasound, which is inscrutable to the novice," sees" the broken bone or the swollen spleen. It does not feel to the expert that there is any inference being made from the pattern on the screen to the nature of the cause of that pattern.

Of course there is a subconscious inductive inference being made, in the sense that the pattern on the screen is evidence which strengthens the radiologist's
belief about the nature of the patient's condition. But that is consistent with the
point made by Ryle and Dreyfus, that in such cases the immediacy of the
judgement depends on the ability, improved by experience, to narrow down the
range of possible interpretations. Novices in a field consider irrelevant hypotheses
where experts recognise situations as falling under a particular concept straight
away. Hence the phenomenon described by Gestaltists as "seeing as".

Unlike some phenomenologists I do not take the gestalt phenomena as
evidence that there are some judgements made which do not require inference:
that would be to take the data of phenomenology at face value. However the
phenomenological data do indicate that that concepts can be applied without
explicit conscious inference and with seeming automaticity. Ryle and Dreyfus are
surely right to suppose that this apparent automaticity cannot be underwritten by
a mind which represents all, or even most, possible interpretations.

Automaticity is the result of the narrowing of possibilities and their
correlation with a narrow range of responses. Ryle and Dreyfus both explain this
contraction as the result of practice and habituation, so that, as Aristotle puts it, the
ability to recognise situations and respond appropriately becomes second
nature. The somatic marker hypothesis is a candidate explanation for the
automaticity of both recognition and response.

At the recognitional stage, the somatic marker hypothesis explains why
some rather than other aspects of the situation are represented as falling under the
relevant concept: because they evoke a characteristic emotional response present
when the concept was acquired. The response which flows from recognising the
situation in this way depends on the representation of possible outcomes,
representations which, as Damasio says, are themselves somatically marked.
Somatic marking ensures that some outcomes rather than others are represented
and retained in working memory during the process of deliberation. Only those
outcomes to which the agent has been habituated or which are similarly
somatically marked are represented. As a result most options are never represented at all, and, of those which are represented, only those which are strongly somatically marked figure in the agent's deliberations.

I asked earlier "Why would the child bother to take an interest in the project of interpreting and predicting her fellows' behaviour in the absence of any feelings of affective engagement?" One answer which suggests itself, and which I dealt with in detail in chapter three, is that, for higher primates at least, the ability to interpret and explain each other's behaviour in intentional terms is advantageous for survival. While this is true, the same applies to affective responses. It must be advantageous to survival if, even from infancy, the primate baby has a tendency to engage with conspecifics, to gain pleasure from contact and feeding rituals with the mother, to respond affectionately to smiles or grooming behaviour and to shrink instinctively from menacing gestures. So it makes evolutionary sense that the emotional dispositions of a primate baby would complement its cognitive abilities.

These evolutionary hypotheses are, of course, not decisive. The case of autism seems to me far more persuasive. In this instance the normal affective responses seem absent from birth with consequent disruption to the ability to apply the relevant intentional concepts. High functioning autistics, whose cognitive abilities are within the normal range learn to apply concepts of belief and desire and to associate particular situations and actions with intentional interpretations but always in a stereotypical and routinised way. I suggest that they are not alive to the intentional possibilities of unfamiliar situations because they lack the phenomenology, the somatic marking, on which the rest of us can rely to determine what others are thinking or feeling in specific situations and to generate appropriate responses. The best an autistic, even a cognitively unimpaired or exceptional one, can hope for, is an alienated understanding of intentional concepts.
Chapter 5  
Cognition in Context  

Chapter 5: Cognitive Anthropology

Part I  

This chapter discusses the argument that the psychology of culture can be explained as the folk psychology of a theory of the social world, and a related claim that the representation of such a theory provides the capacity for theory representation modeled by Tolkoff. Since I deny the Tolkoff view, I adopt the idea that the construction and maintenance of culture depends on developing the capacity of psychological capacity. Tolkoff was preoccupied to explain, I provide an alternative account of culture, extending the argument of constructivism developed in chapter 4, and the relationship to the psychology of culture.

1. Anthropology and Intentional Interpretation. Making the case for treating culture as social in nature, as an agentivist as the ability to induce, or engulf agent action for the understanding of cultural evolution. This is to say that the argument between Tolkoff and anthropologists is not in the understanding of culture as intentional action. We can develop an account of culture, as in the way Tolkoff’s account as in the way ı propose, as developing a theory of mental representations, which do not include unifying symbolic principles.

2. Anthropology and Psychology. The relation between anthropological inquiry, cultural psychology, and cognitive psychology, which requires social cognition to account for cultural representation.

3. Tolkoff’s the Folkology of Cultural Belief. The argument that a Tolkoff’s mindset includes a folk psychology which is the construction of a theory representation. The essential explanation of the universality of cultural belief, which is the representation to form automatic mental simulations. Such simulations can be explained, as in Tolkoff’s account, without involving the postulation of a Tolkoff.
Chapter 6. Cognitive Anthropology

Precis

This chapter discusses the argument that the acquisition of culture can be explained as the internalisation of a theory of the social world, and a related claim that the representation of such a theory exploits the capacity for theory representation postulated by ToMM. Since I deny the ToMM claim, but accept the idea that the acquisition and transmission of culture depends on essentially the same sorts of psychological capacity ToMM was postulated to explain, I provide an alternative account of culture, extending the account of context construction developed in chapters four and five.

6.1. Anthropology and Intentional Interpretation. Makes the case for treating social cognition in adults as an extension of the ability acquired in infancy, to construct a context for the interpretation of behaviour. If this is correct, then the arguments between ToMM and competing explanations of social cognition in infants can be replayed for the understanding of cultural communication.

6.2. Evur amongst the Fang. A specific example of a set of cultural beliefs which can be explained as an internally represented theory, or in the way I prefer, as depending on a disparate set of representations which do not include unifying explanatory principles.

6.3. Anthropology and Psychology. The relation between anthropological theory, which explains actions in terms of sets of beliefs and desires, and cognitive psychology which explains action in terms of mental representations.

6.3.1. ToMM and the Rationality of Cultural Beliefs. The argument that a ToMM module, initialised in childhood, could serve as the representative medium which explains the construction of contexts for cultural communication. The essential explanandum is the rationality of cultural belief, which implies an ability to form second-order mental attitudes. Such attitudes can be explained, as in chapters three and four, without the postulation of a ToMM.
6.4. Bourdieu on Rule Following. The central argument of the thesis, and the conception of practical knowledge on which it relies, restated via a critique of Bourdieu's account of cultural acquisition.

6.5. Anthropology and Psychology Again. Bourdieu's account compared to that given by cognitive anthropologists.

The previous chapter argued that the autistic deficits were not due to malfunction of a module which reads meaning into activity by projecting representations into the heads of other people via a theory of other minds. Rather, the autistic child, due to an affective deficit, fails to bring to bear a complex of abilities we normally rely on in understanding others: face and expression recognition, emotional sensitivity, recognition of typical situations and sequences, together with linguistic knowledge, in order to construct a context for the understanding of another’s words or deeds. Treating intentional attribution as resting on a complex of abilities mobilised in the project of context construction allows us to understand intentional interpretation as a cognitive achievement, without the need to treat it as requiring explicit representation of a theory of other minds.

This chapter looks at context construction in the adult world and examines the same issue: should we explain it in terms of an internally represented theory? The question is interesting because some have argued as follows:

P1. Context construction in the infant and adult worlds is sustained by the same psychological processes and neural architecture.

P2. Context construction in the infant world is sustained by the functioning of a ToMM module.

C. Context construction in the adult world is sustained by the functioning of a ToMM module.
Naturally I disagree with the second premise and, therefore, the conclusion. However in considering some of the reasons which have made this argument attractive I can clarify my reasons for rejecting the ToMM approach and reconnect my account with the considerations raised in chapter two about the acquisition of culture. In that chapter I gave arguments against treating the acquisition of culture as the internalisation of a theory of the social world. The argument above is, of course, very congenial to the idea that cultural knowledge is tacit knowledge of a theory: someone who accepts this argument can say that the young child develops an ability to represent, as an articulated theory, the concepts necessary for context construction. As an adult, the module which represents the theory is applied to the understanding of the wider cultural milieu, but the abilities on which social cognition depends are the same as those originally acquired in infancy. I do not want to dispute the first premise, only the empirical hypothesis about the nature of context construction which underlies the second premise. Before doing so, we need to see the attraction of the idea that the acquisition of culture depends on the internalisation of a theory of the social world.

6.1. Anthropology and Intentional Interpretation

Quine's quasi-anthropological approach to language understanding depicts the child, or language learner, as a stranger who must interpret the words and deeds of an alien culture with which she is confronted. Quine allows observation

---

1 The research on the child's theory of mind shows that this empiricist approach is wrong. We do not need to construct a theory of other minds in order to read meaning into behaviour. The normal child is pretheoretically committed to the notion of other minds and to the idea that communication is meaningful. However Quine might respond that he is interested in giving an account of meaning, *sub specie aeternitatis*, and, as such, he cannot presume the anthropocentric preinterpretation of any data. He must act like an scientist constructing explanations of communicative behaviour: using only data which is equally available to all theorists, not just those for whom it is innately or culturally preinterpreted. The result of such an activity is a *philosophical* theory of meaning. Dummett, and, possibly, Wright are others who have this claim. For a discussion of these sorts of views which contains a hostile response to their verificationist aspect see (McDowell 1981) especially p230.
statements about behavioural patterns of stimulus and response as the minimal basis for the construction of theories about the meanings of the words and gestures with which we are confronted: the neo-Quinean extends from behaviourism to theoretically anchored mentalism.

In support of this approach she can cite social sciences, like anthropology, economics, political science and history which explain people's behaviour by attributing sets of beliefs or desires. These explanations are expressed as theories of a domain of activity (religious observance, decision making under uncertainty, voting behaviour or participation in revolutionary movements). Insofar as these disciplines seek to explain the activities of individuals they cannot ignore the question of how the relevant theory depends on individual psychological processes. For instance an economist who explains behaviour in terms of rational choice models of decision making is often asked whether she is making a claim about the psychology of individuals: that they decide what to do by computing preference rankings, or whether she is merely claiming that rational choice models succeed in predicting aggregate economic behaviour. The response most congenial to the neo-Quinean position is to say that social science explanations describe, in propositional attitude vocabulary, the structure of an internally represented theory which governs behaviour.

For example, anthropologists attribute beliefs in the efficacy of witchcraft in order to explain ritual activities. Or they describe the transmission of 'traditions', regarded as theories which articulate a folk metaphysics or cosmology. An example is Horton's explanation of African tradition as a counterpart of western science: both are theories which postulate unseen entities and forces in order to explain the phenomena observable in everyday life. The observable phenomena are systematised and explained by commonsense notions, first order assumptions, but explained at a deeper level by theories, second order assumptions, which invoke the power of gods or weak nuclear forces. So for Horton a tradition is a
"single overarching framework of secondary theoretical assumptions" shared by participants in the tradition. (Recall that, according to Quine, electrons and "Gods of Homer" are both entities posited by theories whose purpose is the explanation of observable phenomena in terms of deeper, unobservable, divine or natural forces). The important point is that Horton thinks that when Africans see or talk about a ritual or traditional gesture they understand it in terms of the theory.

However, participants in a tradition do not explicitly represent the anthropologists' theory of that tradition: they just remember how to perform singular instances of traditional practices and perform them. If asked, they cannot give an explanation which locates the practice in a deeper theory. Of course the anthropologist may claim, in such a case, that theory is internally represented at a level inaccessible to consciousness: in which case the anthropologist has become a cultural psychoanalyst whose explanations face the typical dilemma of psychoanalysis. Either they are empirical hypotheses about the nature of unconsciously entertained mental representations and a hostage to psychological theory, or they are not such hypotheses but interpretative narratives or hermeneutic strategies insulated from any findings by a future science about the actual psychological processes involved in the causation of behaviour.

Recognising this dilemma, anthropologists like Sperber and Boyer argue for a minimal construal of the implicit psychological claims of anthropology. These anthropologists reject the idea that traditional practice is explained by the internal representation of a set of propositions which comprise an articulated theory amounting to a world view or tradition. Sperber and Boyer argue that the actual representations which underlie cultural behaviour will not include a theory of the domain in question (the supernatural powers of witches for example). Rather they will be the disparate memories, beliefs and other psychological representations

\[^2\] (Horton 1982 p229)
which explain the practice of the individuals who maintain and transmit traditions.

In chapter one I said that theoretical knowledge was knowledge of a set of propositions whose logical relations articulate a particular domain. The epistemic role of theory is to allow belief revision and formation by the evaluation of evidence according to its consistency with the rest of the theory. Sperber and Boyer deny that traditional behaviour is explained by the conscious or unconscious explicit representation of a theory in this sense because the propositions actually represented by the participants in a tradition do not constitute a unifying explanation of a range of phenomena in terms of more fundamental principles. I shall add to their argument a claim that the relevant representations ground practical knowledge of propositions about what to do and say in specific instances rather than intellectual knowledge of principles which subsume those singular instances. This is not to deny that traditional beliefs are attitudes to a set of propositions, however the relevant propositions are known practically and not as the result of intellectual knowledge of an integrated theory.

According to Sperber and Boyer, anthropology should be redirected away from the production of hermeneutic narratives masquerading as theories about the structure of la pensee sauvage, to the study of the way in which the actual representations on which culture depends are produced and transmitted. Alternatively, anthropology should divide into the production of hermeneutic interpretations, not constrained by underlying psychological processes, and a cognitive anthropology which tries to determine the actual relation between psychological processes and cultural practice.

Other anthropologists maintain that the attribution of traditional beliefs can avoid the issue of mental representation by focusing on the behavioural dispositions of the participants in a cultural practice. However, someone who claims that when anthropologists explain other cultures using psychological
language they systematise the dispositions of their informants, needs to say why one, rather than another systematisation is to be preferred.\textsuperscript{3} The hermeneutic account offers no guide here since it is, as Sperber and Boyer point out, psychologically unconstrained. Another solution, implicit in the work of Bourdieu, is to say that the anthropologist’s systematisation can function as a rule for the production of traditional practice in the absence of representation of that rule. In other words, explain the transmission of culture, construed as a rule for coordinating social practice, as an example of rule-following behaviour discussed at 2.3. and 2.4. Bourdieu is right to characterise anthropology as a peripheral theory. But Boyer and Sperber are right to point out that such peripheral theories are applicable only because of the presence of a set of mental representations which underlie the dispositions systematised by anthropology. That set will not, however, include representations of a set of principles unifying and explaining the social world.

\textbf{6.2. Evur amongst the Fang}

The Fang People of Cameroon have a term, \textit{evur}, which they use in contexts for which the concepts and categories of everyday life are insufficient. This does not mean that \textit{evur} is not an everyday topic of conversation, it is. \textit{Evur} is related to witchcraft, magic, to traditional story telling, healing rituals, omens, propitious and unpropitious events. People whose abilities are out of the ordinary are said to be \textit{beyem}, possessors of an \textit{evur} which allows them to do well in business and raise productive plantations, become a shaman or performer of traditional ritual, cast

\textsuperscript{3}This of course is only a problem if two systematisations are equal in predictive and explanatory power. Anthropology and social science in general are full of such instances, of competing interpretations of collective behaviour which seem equally adequate to the phenomena. Roy Ladurie’s explanation of witchcraft beliefs (discussed later in the chapter) is an example of an explanation, which attributes psychological attitudes to the participants in a practice, which seem implausible. Others have sought to explain the same behaviour by attributing as different set of beliefs, such as belief in the literal truth of the bible. (Skinner 1988) The point is that both Skinner’s and Ladurie’s explanations are not prima facie inconsistent with the behavioural evidence.
and retract spells against others who have no evur or whose evur is weaker than theirs. Those who are beyem are in touch with ancestors and have access to the nocturnal spirit world of the forest beyond the boundaries of the settlement. So, although discussion of evur is commonplace, it is differentially embedded in the structure of everyday life to discussion of other everyday topics.

To have evur is not a 'surface' property, so one cannot observe whether or not someone is beyem. External criteria are indirect and insufficient evidence for the possession of evur. Nor is the performance of activities associated with evur a guarantee of its possession. Often a story teller who knows the epics and can perform them accompanying himself on the mvet is not considered a beyem. Similarly, there are people who perform healing rituals yet are not considered a genuine shaman or ngengang. In these cases the knowledge of the epics and of the appropriate ritual is insufficient since it is not accompanied by evur. To have evur is an all or nothing matter, and it cannot be acquired or discarded. Instead it is recognised through its association with typical events such as recovery from serious illness. However these events are not necessarily correlated with evur. Often people admit that they were mistaken, that a person who they took to be beyem was not after all. 4

It is phenomena such as these which invite the "theoretical" approach suggested by Horton. The idea is that concepts like beyem, evur, ngengang, and mbommvet are understood and applied via a coherent set of beliefs which can be represented as a theory of the way in which occult forces are manifest in the observable world. It is consistent with this view that, as Churchland says, observational judgements would be informed by theory. Thus when one Fang sees another singing a traditional epic he sees him as ngengang because the perceptual judgements he makes of such situations are conditioned by the tacit operation of the theory of Fang metaphysics.

4Pascal Boyer has given a full ethnographic description of the Fang in (Boyer 1982) and (Boyer 1988).
There are two, related, objections, to this suggestion. The first is that it is doubtful whether the Fang tribesman who explains another's actions in terms of his possession of *evur* actually possesses, tacitly or explicitly, anything like a theoretical definition. Horton might reply that this is true for the use of many theoretically defined concepts in all societies. There is a linguistic division of labour between those who actually know the truth conditions of theoretical terms, and those who merely use those terms without knowing the details of the relevant theory. When the average person speaks of the HIV virus, of T-cells and AIDS, she has little idea of the full theoretical account which links these concepts together: so she actually has only partial knowledge of the truth conditions of her utterances. Nonetheless we still credit her with thoughts about HIV and AIDS, even though the truth conditions of those thoughts are determined by the satisfaction or otherwise of theories she does not completely understand. This example generalises: very few people would be able to give an adequate definition of many terms in common use, but we do not want say they do not understand their words or that their talk is meaningless. Rather we say that their words are used in deference to a convention that knowledge of their truth conditions is to be determined according to the best theories of those who actually work in the area. The extent to which this phenomenon is true is a matter of controversy, but it is almost the orthodoxy for natural kind terms and others whose truth conditions are theoretically contaminated in a similar way.\(^5\)

---

\(^5\)The philosopher who has made the most of the notion of theoretical kinds is Tyler Burge. He regards the account as valid, not just for natural kind terms, but for artefacts and almost any term which is used in deference to a convention that argument over its extension is to be determined by the community, according to standards administered by a group of experts, be they lexicographers or scientists. Burge's account grows out of an attempt to solve the problems posed for a Fregean account of language by the way in which the context dependence of language use apparently vitiates any attempt to give the meaning of a sentence in terms of a set of context free necessary and sufficient conditions of application understood by each language user. See (Burge 1977; Burge 1979; Burge 1979a; Burge 1986b) For a sophisticated application of the idea to conceptual change see (Field 1977). Steven White has a discussion of the issues and a subtle solution to the problem of determining how much of the meaning of a sentence is dependent on individual psychology and how much is dependant on social and environmental conditions. He treats the former as determining a partial function to the latter. (White 1982)
Chapter 6  Cognitive Anthropology

Perhaps something similar is true of the Fang. The average tribesman does not know all of the arcane lore about *evur*, but the initiated *shaman* and *ngengang* do, in the same way as the geneticists at the Pasteur Institute are presumed to possess the relevant explicit theoretical knowledge about HIV. To be a shaman depends on possession, indeed the monopoly, of the relevant expert knowledge. This sort of response is most congenial to Horton's approach, and it can be backed up by sociological and historical analyses of the way in which specialist monopoly of expertise functions to preserve social structures.

However, and this is the second problem, this sort of analogy between western science and witchcraft depends on the truth of the presumption that the distinctive role of the *shaman* is guaranteed by his possession of knowledge which can be plausibly represented as a theory of Fang metaphysics. The original idea was that we could regard the Fang as believers in a metaphysical theory of occult power, even though they did not mentally represent that theory in forming their judgements about its manifestations, *provided* that those judgements were like our judgements about HIV: i.e. tacitly containing a reference to the experts who do understand the relevant theory.

However, discourse about *evur* is not regulated at any level by knowledge of a set of principles which explain instances of *evur* manifestation in terms of its essential, or even theoretically-postulated accidental, properties. Boyer has identified three levels of discourse in which *evur* and its manifestations are discussed.6 The first is common discourse, in which events, such as murder, known to manifest *evur* are discussed. In these cases the imprecise folk wisdom concerning *evur* is invoked. (Such as the knowledge that when an *evur* bearer goes to sleep, she in fact goes to *mgbel*, the forest village where all witches meet to fight. These fights produce illness as a person's *evur* is weakened in the fighting.) While these conversations show a great deal of shared knowledge about *evur* they do not

---

6My account is a precis of chapter 2 of (Boyer 1991)
purport to state anything definite, and certainly nothing explanatory. This is what I've been told, but after all I'm not an ngengang, a typical remark cited by Boyer.

A second context in which evur is discussed is private gossip. Unlike the common discourse, gossip does not deal in vague generalisations subject to infinite qualification by putative experts. Rather, gossip retails singular instances of the manifestation of evur. For instance someone has heard that a specific ritual has been performed, although they did not see it themselves. Gossip is distrusted and instantly dismissed as 'lies' or 'empty talk' on the basis of its hearsay nature, as opposed to personal testimony which can be argued against on grounds of its content. Gossip can thus not claim to state any general truth about the nature of evur because it is automatically suspect as hearsay, produced in a context likely to be contaminated by personal interests, and restricted to reports of singular events rather than explanation of those events.

The third discourse register is 'the expert context' in which shaman, witch doctors, healers and diviners, those who possess evur, make definite statements about the cases in which evur is manifest. For instance divinations are often organised after deaths. On these occasions the expert presages his remarks with general statements which rehearse the commonplace knowledge, such as "those who engage in things of the forest eventually must pay for it" before giving his explanation of the death. It is here, in these explanations of singular occurrences, that one might hope to see evidence of the theoretical knowledge which underpins the possession of expertise. Boyer cites the case of a boy who died mysteriously. The shaman diagnosed the case as follows: the onset of the illness came after the boy's attendance at a performance of the epic songs. Everyone should recall that there were many ants in the house where the performance took place and that the performer had stuck his tongue out during the performance. This was a satisfactory explanation since ants are associated with witchcraft and, for the Fang, sticking one's tongue out is a mysterious and sinister gesture. The final
diagnosis was that the boy's *evur* was strong and malevolent and could not stand the presence of another *evur* bearer, the epic singer.

The shaman's explanations thus do not depend on knowledge of a theory of the nature of *evur* any more than its discussion in the other two discourse registers. Rather they draw together salient events which are recognised in the common discourse about *evur* to be manifestations thereof, without providing any deeper explanation of the relation between those events.

This lack of generalisation is not a strategy of concealment by the shaman, an attempt to retain his prestige by refusing to disseminate his knowledge. Rather it indicates that discourse about *evur* is not regulated at any level by the possession of a theoretical definition of the term. This is seen even more clearly when we look at the way shaman, or story tellers, are initiated. There is no process of instruction in the nature of *evur* and its causal powers. Rather there is instruction in the specific causal powers of certain herbs and medicines and exposure in the presence of other shaman, to specific scenes of *evur* manifestation. The end process is that the initiated becomes *ngengang* or *mbommvet*. Accounts by initiates usually explain their abilities in terms of specific acts and performances they have undergone, together with access to the spirit world, normally the result of a mysterious and unmentionable episode, (typically the purchase of their outstanding capacities for the life of a parent.) The point is that encounters of this nature, rather than a theoretical articulation of them in terms of underlying properties, are sufficient for expert knowledge of *evur*. "What makes certain people able to identify the effects of *evur* is a series of experiences. The expert does not acquire another, more refined 'definition' or 'characterisation' of *evur*; he or she acquires a repertoire of salient memories, which concern singular situations not abstract principles."7

Boyer wants to remind us that tradition is best explained in terms of concrete rituals, the recognition of specific and singular situations, literalism and

---

7(Boyer 1991p36)
repetition, rather than the transmission of a "world view" considered as an articulated set of propositions which state unifying principles. Transmission of beliefs about *evur* does not depend on tacit representation of a "Fang cosmology." Rather it requires the repetition of ritual, the memory of specific instances of its manifestation and the interaction of levels of discourse which preserve its central social importance.

Boyer dismisses the notion that all these specific performances and mental events are united by the tacit application of a metaphysical theory. What the Fang bring to the discussion of *evur* is not the mutual possession of a definition, or the tacit agreement to have their discourse regulated by those who do, but memories of concrete situations and events. Each situation is not deductively subsumed under a shared theory but assimilated into memory by a process of non-demonstrative inference. The psychological structures which underpin these non-demonstrative inferences are the focus of cognitive anthropology.

### 6.3. Anthropology and Psychology

The project of cognitive anthropology is thus, firstly, to discover the nature of the concepts represented in the minds of those who participate in a tradition. And, secondly, to show how those concepts are acquired and transmitted without relying on hypothetico-deductive inference. The view of culture being rejected is that of the acquisition through socialisation of a theoretically articulated "world view" which is then tacitly transmitted by cultural interaction. Rather, the cultural anthropologist determines which objects, artefacts, practices and situations are salient within the culture and tries to account for their repetition in terms of the psychological processes which underlie the communicative acts which play a role in repetition. The exact nature of these psychological processes is still elusive, but cognitive anthropologists are naturally drawn to accounts of propositional
attitudes which do not explain them in terms of an internalised theory. The types of structure which could serve as a basis for cognitive anthropology are a variety of modular capacities to structure information. Candidates are mental models, prototypes, scripts, concepts of animate and inanimate causation, biological and artefactual kinds, and persons as the authors of intentional actions (the subject of the previous two chapters).

The point, which is very clearly made by Boyer, is that one’s cultural knowledge is sustained by a variety of different cognitive processes whose output is a set of representations which make up that knowledge. These representations are not the result of the application of a single theory-driven inference engine, but the result of many different ways of learning and remembering the information on which cultural repetition and innovation depend. His objection to the theory model is not so much that it postulates a set of propositions which are known by those who transmit and sustain a culture, because he would allow that the dispositions of a cultural participant can constitute practical knowledge. Rather he objects to the spurious unification and integration of the disparate practices and representations which make up a culture in terms of knowledge of an integrated theory.

Boyer’s idea is that concepts become culturally salient, not by being subsumed under a theory which encodes a "world view" but through the operation, within individual minds, of modular capacities. An example is the notion of a script. Some developmental psychologists propose that an important element in infant development is the ability to recognise and generalise from familiar narratives of everyday life. So children in North America, where most of the research has been done, can recognise and reconstruct typical breakfast routines for example. The script can be decomposed into constituent structures which govern whether or not a series of actions fits the script. Some elements, such as the identities of the actors or the foods consumed, are interchangeable, and
some are not: for example, only breakfast foods can play a role in the breakfast script. The central idea is that the concept of a script functions like a template which can be inscribed with a particular sequence and content. The initial inscription sets limits on what can be subsequently understood in terms of that particular script, allowing only certain variations. Fivush has proposed, persuasively, that early pretence play in childhood is governed by scripts. Games of object substitution depend on the recognition of an identity of functional role, defined by a script, between the object and its substitute.8

If this is correct, then traditional practice may be exploiting just such a psychological capacity. The emphasis on literal repetition and the constraints on variation of traditional stories and practices, may both be explained in terms of the way a template is initialised in the minds of participants. The cultural salience of certain practices or situations may be explained, not in terms of theory-driven inferential links between beliefs about those practices and other traditional beliefs, but in terms of their role in a narrative structure. So when the Fang witch doctor explains the boy’s death by citing the ants in the hut on the day of the epic performance, his audience is satisfied, not because he has given a piece of evidence neatly subsumed by their metaphysical theory, but because he has constructed a narrative of events which fits the appropriate template.

Other culturally salient concepts may depend on different modules. For example, there is evidence that very young children are sensitive to, and rely on, a difference between property and natural kind terms. Children are natural "essentialists" rather than description theorists in their understanding of the biological world.9 They classify objects in terms of underlying essences and kind

---

8(Fivush 1987)
9(Gelman 1987) See also Keil who reports similar interesting experiments designed to test children's development of natural kind concepts in the biological domain. He reports a shift in the way similarity judgements are formed from classification in terms of surface similarity (diagnostic features) to classification in terms of shared underlying properties (defining features) or "hidden essences". (Keil 1989)
membership rather than surface similarity. In one experiment four year olds are shown pictures of dolphins and tiny tropical fish and told that dolphins breathe above water and the tropical fish below. They are then shown a picture of a shark, which closely resembles a dolphin and asked if it is more like the tropical fish, to which it bears little superficial resemblance, or the dolphin. The children say that the shark is more like the fish than the dolphin. It turns out that children's similarity judgements typically classify organisms in terms of fundamental, underlying, properties rather than surface resemblance. Similarly, children are more likely to make these sorts of inferences, treating animals as members of a species previously identified by a salient exemplar, where the relevant properties are more 'inherent', such as breathing or feeding, than in superficial cases such as size or speed of movement.

Boyer, extending from examples such as these, proposes that notions such as beyem and evur are understood via the initialisation of an essentialising template. As he says "a single use of a term results in spontaneous ontological hypotheses, which by themselves are the basis of predicate restriction." So children told that certain people are ngengang or beyem because of their possession of evur acquire the concept of an essential division between those who have and do not have evur. This is why subsequent singular performances are assimilated into one or other category, rather than treated as evidence to be subsumed under a general theory. This applies across all discourse registers. The experts, who rely on memories acquired during their initiation, indicate which singular instances are salient, and the rest of the people make an inductive generalisation exploiting the "implicit idea

---

10Sperber surmise that racism may depend on the cultural initialisation of this module. "the initialisation of an ad hoc template for racial classification could well be the effect of parasitic cultural input information on the higher level learning module, the function of which is to generate ad hoc templates for genuine living kinds domains such as zoology and botany. If this is correct - mind you I am not claiming that it is merely that it may be, then no racist disposition has been selected for." (Sperber 1995)

11(Boyer 1993)
that all *beyem* belong to a natural kind*. No one recognises *evur* by applying a definition.

These remarks are speculative but they suggest an explanation of traditional belief as a resting on a *melange* of modular capacities for concept formation and application. They suggest that ritual performance is understood via scripts and the concept of *evur* via the initialisation of an essentialising template.\(^\text{12}\) If this is the case then explaining Fang tradition in terms of a set of coherent beliefs, integrated as a theory and 'passed on' from generation to generation is impossible. There simply is no such theory.

This attack on the notion of tradition as an attitude of belief toward a shared set of propositions is congenial to the Rylean way of understanding mental states. What is it to believe in *evur*? It is to be afraid to go into the forest after dark, to participate in rituals, rely on divination, to gossip about death and illness and so on. It is an open ended pattern of dispositions to exercise one's executive and imaginative capacities according to the appropriate cultural constraints. Sperber and Boyer are, of course, not content with the dispositional account because they want to explain the cognitive processes on which a particular dispositional profile rests. According to Sperber, in particular, to adopt the dispositional account is to abstract away from scientific psychology, which is essential in order to find out why we have the dispositions we do. As he says "We might be tempted to say then, as many philosophers have, e.g. Ryle (1949) that a belief is a disposition to assent to a proposition. As psychologists, however, we will want to go deeper: what kind of mental states might bring about such a disposition?"\(^\text{13}\)

\(^\text{12}\) Often theorists like Keil take their experiments to show that children are "little scientists " unifying surface phenomena in terms of shared essential features. Science is an extension of this process. The important difference is in the employment by scientists of articulated explanatory hypotheses to explain the relation of the underlying properties to the surface features. So while I agree that some beliefs, especially those involved in biological taxonomy are acquired via the initialising of an essentialising template, this does not undermine my argument that culture is not a matter of representation of a coherent theory of a particular domain. My point is that some of the knowledge on which culture depends may be acquired via such a template but all the knowledge on which culture depends is not unified via its subsumption under an explicitly represented theory.\(^\text{13}\)(Sperber 1990p32)
Ryle, of course, disputed that as a psychologist one should want to go deeper. He would agree that the causal transactions between a human brain and its environment are of interest to physiologists and neuroscientists, but that such information was not necessary in order to give a materialist account of propositional attitude psychology. As we saw in chapter two, Ryle accounted for ascription of propositional attitudes in terms of the display of the logical relations among the dispositions of an agent, irrespective of the structure of the categorical base (brain or mental processes) of the dispositions. Along the way he rejected the view that the logical relationships among the dispositions causally depended on the internal representation of the peripheral theory (propositional attitude psychology) which systematises them.

6.3.1. ToMM and the Rationality of Cultural Beliefs

Sperber and Boyer dismiss this view as well, arguing that a variety of psychological mechanisms may be responsible for the initialising of cultural templates. Sperber gives many examples of the acquisition of culturally salient beliefs via modular first order capacities. However he then goes on to say:

An organism endowed with perceptual and first order conceptual modules has beliefs delivered by these modules, but has no beliefs about beliefs, either its own or those of others, and no reflexive attitudes to them. The vocabulary of its beliefs is limited to the output vocabulary of its modules and it cannot conceive or adopt a new concept, nor criticise or reject old ones. An organism endowed with a meta representational module can represent concepts and beliefs qua concepts and beliefs, and evaluate them critically, accept them or reject them on metarepresentational grounds. It may form representations of concepts and of beliefs pertaining to all conceptual domains, of a kind that the modules specialised in those domains might be unable to form on their own, or even to
incorporate. In doing so, however, the better endowed organism is merely using its metarepresentational module within the module's own domain, i.e., metarepresentations.  

Sperber's is an elegant model of the type of mind needed to transmit culture. The mind here is pictured as involving three tiers: a single thick layer of input modules, just as Fodor says, then a complex network of first order conceptual modules of all kinds, then a second order metarepresentational module. 

Sperber finds this model congenial because he is interested in the role rationality plays in the transmission of cultural beliefs. Human agents do not passively reflect their cultural environment, they use the knowledge they acquire therein to deliberate and reflect before acting in that environment to change or preserve it. They question of why agents have the beliefs they have, why some beliefs are retained and others discarded, cannot be answered, without, as Sperber points out, raising the question of how humans think about their own and others thoughts.

Sperber is referring to the way our beliefs about what others are thinking, or what we ourselves should do in a particular situation, can be strengthened or weakened by evidence. To accept or reject such evidence requires that we form beliefs about first order mental states, in other words to consider whether our beliefs are strengthened or weakened by the new evidence. If this is correct then metarepresentation is not just a label for the ability to recognise and respond to other's mental states, but looks, at first glance, like a reinstatement of the theoretical hypothesis at the second order level. How else can the "acceptance or rejection of first order beliefs on metarepresentational grounds" be construed?

If a first order thought is a representation, (or parallel array of representations) then thinking about thinking is a metarepresentation. In this minimal respect it must be true to say that humans can metarepresent. Note,

---

14 (Sperber 1995 p38)
15 (Sperber 1995 p41)
however that in this sense of metarepresentation the musician in chapter two who
learns her part in the string quartet is a metarepresenter. She revises her actions,
bringing them into line with those of her with her colleagues by trial and error.
Thus, in the Rylean sense, she has beliefs about her first order beliefs. In fact this
was the point of the discussion of rule following in chapter two. To be rational is to
be able to revise beliefs according to some principles for weighing evidence and
combining concepts coherently. The initial challenge to the dispositional account of
beliefs was that it could not explain this type of rationality without postulating
some inner representation of a rule or principle for belief revision. The example of
the musician was of someone who does adjust their behaviour in a rational way,
but without explicitly representing a theory of the relevant domain in order to do so.

Of course even if we grant that the musician does succeed in being rational
without explicitly representing a theory in the sense rejected by Ryle and Dreyfus
the question remains: on what type of representations does this type of second
order thinking depend? Ryle of course dismissed this question as irrelevant since
he thought that second order thinking, like first order thinking, was just a matter
of having certain dispositions. Sperber, however, while he might grant, for the sake
of argument that everyday talk is about dispositions, still thinks there is an
interesting question in psychology, cognitive anthropology or cognitive science,
about the nature of the representations on which the dispositions which invite the
attribution of second order thought, and hence rationality, depend.

Sperber has proposed that this type of second order thinking depends on
the "other minds" module. The idea is appealing for at least two reasons: the first
is the intuition that thought precedes language, even if we grant that language
gives us a medium for its articulation which gives the deliberation of language
using creatures some distinctive properties. So we want to say that chimpanzees,
for example, can have a limited range of metathoughts despite their lack of natural
language. Secondly, it seems the child is innately equipped to develop beliefs about others, and its own (recall the change of representation experiments at 3.3.3.) mental states. Sperber suggests that the assimilation of cultural representations is a result of the application of the same mental capacities the child uses to interpret others' words and actions: both involve the formation of beliefs about others' states of mind.

Like Sperber, I believe that the abilities acquired in infancy to interpret behaviour as meaningful are continuous with those we deploy in interpreting the range of culturally loaded information with which we are presented. However I also argued that there is no "other minds module." Rather, I suggested, we focus a battery of capacities in order to represent the psychological elements of a context for the interpretation of behaviour. This position is independent of the issue between Ryleans and neo-Quineans over the referents of mental state terms. I am inclined to agree with Ryle that ordinary practice has more to do with systematising dispositions than stating hypotheses about the nature of inner states. However, whether or not that is true, the issue considered here and in the previous chapter, is the nature of the mental representations which underlie those dispositions, the question in which Sperber is interested.

In this sense, it is clear that an agent believes, desires, expects or fears that \( p \) if the state of her brain represents a certain world, actual or possible. The same is true for having beliefs about beliefs. Someone counts as representing a second order thought if their brain represents a world consistent with the content of the proposition to which the relevant metabelief is an attitude. I have tried to show that the sorts of abilities which invite the ascription of metabeliefs, chimpanzee deception, and the children's passing of the false belief test, can be explained without invoking a modular capacity which employs a theory in order to hypothesise about the nature of representations entertained in the minds of others. The neural circuitry is connected in an elaborate system of feedforward and
feedback loops depending on a variety of systems which are only ever partially integrated. There is no site at which the output of all the relevant circuitry is integrated and which controls the process.

If this is correct, then an appropriately supplemented dispositional account may start to seem attractive. What is it to believe *euur* is efficacious? It is to be afraid to go into the forest after dark, to participate in rituals, rely on divination, to gossip about death and illness and so on. Each of these dispositions, of course, rests on capacities which are the province of psychology. So the correct anthropological explanation is the one which systematises cultural dispositions, as a peripheral theory, and points us towards the mental representations which explain them. My disagreement with Sperber is thus empirical, over the interpretation of the evidence for a modular capacity for mindreading.

### 6.4. Bourdieu on Rule Following

Chapter two discussed the musician who learns her part without explicitly representing the score. In such a case it is appropriate to describe the musician as having second order attitudes, for example it is true to say of her that she wants to come to have true beliefs about the way her part of the quartet is played. Or when she corrects herself that she believes her former beliefs about her part were wrong. So learning requires the formation of metabeliefs in the sense that the relevant embedded propositions are represented in the brain of the musician, or anyone else who does some principled belief revision. However, as I pointed out in the case of the musician, the ascription of those sort of second order propositional attitudes does not commit us to any particular hypothesis about the way they are realised.

The discussion of anthropology suggests that what the anthropologist does is articulate the contextual dimension of cultural communication by describing the
propositions, including those which capture facts about mental states, which are known practically by participants in a culture. That project does not commit us to any particular view of the way the relevant propositions are represented.

Pierre Bourdieu is an anthropologist who conceives of anthropological knowledge as knowledge of rules which function as cultural norms, knowledge of which is required for the construction of a context for the communication of cultural information. Bourdieu denies, however that the acquisition or transmission of those rules requires their representation in the minds of agents. Hence his argument with Jon Elster over the theoretical status of the epistemology of rational choice social theories. Elster's political sociology shows how social change can be explained in terms of individual choices made under certain constraints, modelled in decision theoretic terms. Elster treats these models as idealised descriptions of individual psychologies, of the thoughts of individuals who make up the larger society. Bourdieu, however, while he acknowledges that these economic models are predictively accurate and truth evaluable, denies that individual agents apply any such models. Rather, he says social interaction is produced as if agents coordinated their actions by applying these rules.

This invites the interpretation of Bourdieu as an instrumentalist about social theory, someone who thinks of the postulates of social science as heuristic devices, literally false of individuals but useful in the prediction of aggregate individual behaviour. However his insistence that individuals actually learn and apply these rules indicates that such an instrumentalist reading is wrong. Bourdieu's theoretical exposition shows that he relies on a concept of practical knowledge akin to that of Ryle and Heidegger. However his arguments are mainly negative, directed against the conception of culture as the transmission of an implicitly understood theory, rather than a positive explanation of the processes by which knowledge is acquired. He tends to explain cultural knowledge in terms of the environment, social and physical and pays little attention to the way the
environment is represented in the minds of individuals and the causal connections of those representations with the dispositions of agents. Bourdieu, like Ryle, explains social understanding as a set of skills, treating these skills as primitive and not requiring further explication in terms of the representations which subserve them.

However, as Sperber points out, the physical and social environment explain culture because individual representations thereof create a context for cultural communication. It is not just the environment, but the environment as represented in the process of context construction, which is relevant to the peripheral theories of anthropology.

Bourdieu, however, proceeds as if there are only two possible accounts of intentional psychology. A dispositional account of intentional states which abstracts entirely from actual psychological details, and a conception of mental states as attitudes to an explicitly represented theory. Aware of the difficulties with the latter account he endorses the former, but in a way which insulates it from psychology, and hence from cognitive anthropology. Even his endorsement of the dispositional account, which focuses on the sociological preconditions for the acquisition of the required dispositional profile, makes it slightly mysterious as to how it is that the resultant profile counts as the following of a determinate rule. Sperber’s account has the virtue of directing us to the mental representations on which rule-governed behaviour depends, while reminding us that the transmission of culture is unlikely to depend on representation of rules considered as theories of the social world.

Bourdieu discusses rules and norms of thought at several places. He objects to a conception of culture as a rule for the production of behaviour, implicit in the anthropology of Radcliffe Brown. Radcliffe Brown treats culture as an abstraction, a Platonic entity of which actual practice is always an approximation. On this view any action is an execution of a model, either a Platonic entity, or a norm
"consciously posited and respected by agents "16 (the agent's mental representation of the model) Neither conception of rules is acceptable to Bourdieu for reasons familiar from chapter two.

Firstly, actions have meaning for agents, and the mere fact that an action is in accordance with a model does not make it meaningful, whether or not the model is internally represented. Secondly, the frame problem. The rationality of intentional behaviour is a matter of selecting the appropriate action from the range of possibilities open to an agent. This range is narrowed down by prior participation in a form of life, not via the deductive application of a theory of the social world. As Bourdieu puts it, lived experience is vital to social understanding, not because rule following is a matter of the coincidence of subjectivities, but because the subjectivity of someone who acquires their dispositional profile against a certain background has "a feel for the social game which makes it possible to take for granted the meaning objectified in institutions." 17This is why a rule exists, not as something to be consulted," but in a practical state in agent's practice and not in their consciousness or rather their discourse "18

Bourdieu here describes a central feature of the Dreyfusian account of intentional activity: the way experience constrains the possibilities for interpretation of phenomena and, hence, of action. The individual must be able to rely on her own experience in coordinating her practices with those of the community. However her experience inclines her to act appropriately because her inclinations are conditioned by previous experience. Thus she "follows the rule" because, in unfamiliar situations, she is inclined to do one thing rather than another. That inclination, which depends on prior habituation, is not produced by a theory derived from past instances and extended to the present case.

16(Bourdieu 1990 p38)
17(Bourdieu 1990 p27)
18(Bourdieu 1977 p27)
Bourdieu thus rejects, for reasons of empirical sociology, a conception of "rules as rails", as timeless concepts which determine every instance of their application, also rejected by Wittgenstein for different reasons. Bourdieu emphasises that social practice is adaptive and flexible. The dispositions which were acquired against one background count as knowledge of the same norm which governs behaviour even when the background conditions change or are disrupted in such a way that practice is inevitably modified. An example might be of a nomadic community which maintains its tradition through war, colonisation, environmental change and resettlement. The material and political conditions of life against which succeeding generations acquire their dispositions will change, but the way in which those material and political conditions are interpreted depends on the dispositions acquired and shared in previous conditions. A new set of interpretations, inculcated dispositionally, then constrains the possibilities for future interpretations and action.

Bourdieu's principal concern is to show how these interpretations can be constructed and transmitted without being explicitly represented and learnt as a theory of the social world. As I said, his arguments are mainly negative, directed against the idea that culture is transmitted by the explicit representation of a world view. His positive arguments are less convincing, since they presuppose, but do not actually argue for, a dispositional account of rule following.

One reason his positive arguments take this form is that they are situated in a theoretical debate between phenomenological and structural accounts of the understanding of social norms. The phenomenologist argues that individual experience is primary in the explanation of action and thus needs to account for the way in which that limited experience can reflect facts about the wider society, such as its class or institutional structure, which seem to be discernible as patterns from the perspective of the social theorist. Dreyfus gives one possible answer: the individual's experience is constrained in specific ways by the structure of the social
world. The structuralist answer is that the individual's experience is not relevant in the explanation of action. Patterns, such as class or economic transitions, discernible at a higher level are sufficient explanations of the production of social structures through individual actions.

Bourdieu claims to reject both these approaches, substituting his conception of practically understood rules which "supply the mediating function between objective structures and actors' conceptions of their interest". However, what Bourdieu really does is outline a conceptual space, a third possibility "to escape from the philosophy of the subject without doing away with the agent, as well as from under the philosophy of the structure but without forgetting to take into account the effects it wields on and through the agent" 19 His articulation of this third possibility is, however, closer to the structural account. He concentrates on a description of the actor's social situation and attempts to show how, irrespective of first person disclaimers to the contrary, an actor's life trajectory conforms to social regularities such as class or educational background. Effectively he articulates the context, using sociological tools, and demonstrates that context constrains actor's possibilities for action. The demonstration consists in showing that actors conform to predictions about their life trajectory which can be made purely on the basis of observations about their material, social circumstances.

The structures constitutive of a certain type of environment ... produce *habitus*, systems of transposable durable dispositions... principles of the generation and structuring of practices and representations which can be objectively "regulated" and "regular" without in any way being the product of obedience to rules, objectively adapted to their goals without presupposing a conscious aiming at ends, or even express mastery of the operations necessary to attain them, and, being all this, collectively orchestrated without being the product of the orchestrating action of a conductor 20

19(Bourdieu 1992 p121)  
20(Bourdieu 1992p72)
Such an account presupposes that conformity to social regularities, transmitted by tradition, education or occupational specialisation can be produced without explicit representation of the context by individuals.

Bourdieu's anthropology actually stands in need of a dispositional account of rule following. Bourdieu provides an explanation of individual action by observing and interpreting patterns of social activity and showing how the individual's actions conform to those patterns. However, as he acknowledges, that will only serve to explain the individual's behaviour if the individual's actions can be seen as an instance of following a rule which brings about that pattern. Having ruled out the idea that individuals represent that pattern mentally and treat it as a goal in the production of behaviour, Bourdieu is left with only the dispositional account of rule following to fall back on.

In chapter two I argued that we could have practical knowledge of rules or propositions without explicitly representing them, in virtue of a pattern of dispositions. However, even if Bourdieu wanted to avail himself of the conception of practical knowledge articulated there, it would not serve his purpose entirely. For that he would need a dispositional solution to the rule following problem which treated dispositions as "bare dispositions" unconstrained by any causal relations to structures in the categorical base. The reason is that, for Bourdieu, no information about the nature of representations actually entertained in the mind of the agent is relevant to the explanation of action. He treats the mind as passively reflecting institutional and social structures with which it is confronted. As a result he explains action in terms of external constraints on action rather than internal constraints on representation.

If Sperber and Boyer are right, then details of individual psychology strongly constrain anthropological theory: because the correct anthropological theory is the one which systematises dispositions which depend on actual psychological processes. I have agreed with this position while endorsing a
different account of the actual processes involved in constructing the psychological elements of context.

Bourdieu is adamant that his account is an explanation of how individuals' actions reproduce culture, so his interpretations of culture are not merely hermeneutic narratives, but accounts of individuals' contribution to the production of social structures. However, while Bourdieu embraces some elements of the dispositional account advanced by Ryle in his references to a "feel for the social game" and his talk about the *habitus* as "embodied history," he never actually explains how it is that the dispositions of individuals can count as the following of a rule for the recognition, repetition and transmission of cultural contexts.

The reason is that his target is a purely phenomenological account of cultural acquisition and transmission. Consequently he conceives of individual psychology as an account of agents' conscious intentions and conscious self representations. And, since the patterns to which agents ultimately conform are not the consciously self-represented goals of the agents themselves, a phenomenological individualist psychology has little to contribute to anthropology. Bourdieu wants to explain "conformity in the absence of the intention to conform." 21 Thus he concentrates on the objectively quantifiable structures which are the outcomes of individual actions and presumes the truth of something like a dispositional account of rule following in order to account for their production and reproduction. Sperber and Boyer are equally unconcerned with phenomenology but they provide an account of the acquisition of and repetition of social structure without abstracting away from the details of individualist psychology.

The advantage of their account is, of course, that it places constraints on the types of theory which can be legitimately advanced by social theorists. If social theory consists merely in providing theories which are consistent with the

21 (Bourdieu 1977p29)
dispositions of members of social groups then there will be no way to choose between competing accounts which "fit" the relevant sets of dispositions.

Consider Le Roy Ladurie's famous explanation of witchcraft beliefs offered in the *Peasants of Languedoc* 22 According to Ladurie the collapse of the social reforms associated with the Reformation foreclosed the possibility of improvement in the material conditions of life for the peasants. Frustration at the halt of social reform was transmuted into a desire for symbolic escape from their misery, which expressed itself in the "chimerical and fantastic revolt of the witches' Sabbath, an attempt at demonic forms of escape." The objection to this explanation is not merely that no peasant would have explained his actions in this way (because after all he might have represented Ladurie's theory unconsciously) but that it is highly unlikely that any representation, conscious or not, entertained by the peasant as a result of his communicative interactions, represented even a component of this theory.

So, while Ladurie proposes a set of mental states which which systematise peasants dispositions, that alone is not sufficient. We reject his account because it seems highly unlikely that representations actually entertained in the minds of peasants peasants could have amounted to practical or intellectual knowlege of a set of propositions about the failure of land reform. This does not imply that we tacitly endorse an account of propositional attitudes as hypotheses about the nature of internal representations, just that we expect our explanations of behaviour to respect the causal constraints on production of behaviour. In other words, we cannot attribute beliefs which cannot be entertained by the mind with which we are confronted in virtue of its causal history and the operation of its representative capacities.

### 6.5. Psychology and Anthropology Again

22(Le Roy Ladurie 1974 p203ff)
The anthropologist explains other people's actions by attributing beliefs, desires, hopes and expectations. In this she is no different from anyone else. Psychologists and economists do the same, and so do children and chimpanzees. We speak of these propositional attitudes as states of a mind. So a natural explanation of someone's actions is their state of mind — the complex of mental states they entertain. Propositional attitude psychology articulates the relations of these states of mind in terms of their inferential relations. It treats them as attitudes to propositions. So a mind is something which can sustain articulation in terms of an inferential network: a theory.

Although this theory is well entrenched in human culture and apparently rests on some innate capacities for communication with conspecifics its details remain obscure. Philosophers have attempted to provide a full account by answering the question: to what do mental state terms refer? This thesis takes that project at face value and evaluates one version of it. This version suggests that theoretical terms in general, and propositional attitude terms in particular, refer to entities whose properties are postulated by hypothetico-deductive empirical theories. Propositional attitude psychology states covering laws which serve as the first premise in a deductive inference whose conclusion is the nature of the relevant mental state: a state of the brain with propositional structure or a mentalese transformation of propositional structure.

If this is correct then the theory of propositional attitude psychology is something we learn and exploit in interpreting other people's behaviour. Either humans are born with the theory of propositional attitude psychology, or a precursor version of it, or they develop it as they are confronted with the problem of coordinating their activities and communicating with others of their species. If it could be shown that this was the case then the "theory theory" version of propositional attitude psychology would have straightforward empirical
confirmation. However, while there is evidence that children do rely on some innate capacities for recognition of conspecifics and interpretation of their behaviour, it does not seem that these abilities are integrated by a specific module devoted to the formation of hypotheses about the connection between others' internal states and their behaviour. Rather it seems that, in virtue of an innate disposition to engage and communicate with others, children direct a range of capacities to the construction of what Sperber and Wilson call a context for communication.

Another account of propositional attitude psychology treats it not as a protoscientific theory, but as a way of systematising behavioural dispositions. The dispositional account of propositional attitude psychology has the advantage of directing us to the range of capacities we integrate in the task of context construction. However, I also claim that the correct dispositional account is the one which is constrained by the psychological processes dealt with by developmental psychologists and cognitive anthropologists. We read meaning into behaviour by classifying the way others' dispositions are connected, using the inferential structure of language to construct an appropriate network. But that classification depends on a series of perceptual and cognitive capacities realised in the structure of the brain.

What then is the nature of the connection between brain structures and propositional attitude psychology, if propositional attitude psychology is not a hypothesis about the nature of those structures? Consider Bourdieu and Boyer again. For Bourdieu a belief in *eur* is a set of dispositions, for Boyer it is a representation in an essentialising module which classifies objects into the categories of natural and supernatural. My claim is that belief in *eur* is attributed to explain a set of dispositions which depend in part on the presence of that representation in the essentialising module, as well as on the presence representations in other modules, perhaps scripts which constrain the recognition
of appropriate actions in divination contexts, and so on. This conception allows that the same set of dispositions could be differently realised by a different set of representations in different modules, but this seems unobjectionable. After all, each of us participates in social interaction by constructing a context according to our different experience, which implies, for example that my belief in *evur* may causally depend mainly on scripts activated through participation in divination rituals and yours may depend mainly on an essentialising prototype activated through gossip and conversation.

Does this insulate anthropology from psychology, or everyday propositional attitude ascription from the scientific investigation of the way the brain processes information? I do not think so. What it implies is that the ascription of mental states in everyday communication does not carry with it a commitment to a particular hypothesis about the psychological processes by which they are realised. Nonetheless such ascriptions are used with the implicit commitment that they be realised. That is the objection to pure structural accounts of human psychology and to the structural elements of Bourdieu's program: it is not sufficient to explain the reproduction of social structures (widely distributed resemblances in patterns of dispositions) by citing only patterns of dispositions and observing that individuals exemplify them.
Conclusion

The discussion of the relation between anthropology and psychology allows me to summarise the central claim of this thesis: we attribute others with beliefs and desires in order to interpret and explain their behaviour, but in so doing we are not stating any hypotheses about the nature of the brain states which cause that behaviour.

If our use of mentalistic vocabulary expresses inductive generalisations from behavioural evidence, without expressing any hypotheses about the causal origins of behaviour, then it looks, at first glance, as if we must be behaviourists, or hold the "bare" dispositional theory of mental states. One obvious objection to such theses is the fact that we treat as candidates for intentional explanation behaviours which can be understood as the product of conscious deliberation or, if not of conscious deliberation, at least the rational adaptation of behaviour to the environment, physical or social. A reductive behaviourism or bare dispositional account of mental states seems to leave no room to articulate the difference between rational and nonrational action.

The attraction of the "theory of mind" hypothesis is that it explains the interpretation of behaviour in terms of the rational relations among mental states by positing a mental module, which develops in stages through early childhood, dedicated to the task of forming higher order thoughts. This module is employed in deliberation, thinking about one's own mental states, and in inferring the nature of others' mental states from the evidence of their overt behaviour. We do this, according to ToMM theorists, by applying a theory of other minds to the behavioural evidence. The theory of other minds arrived at by four year olds is the same as that employed by adults. It treats mental states as logically opaque representations of propositions which cause behaviour. ToMM is thus a non behaviourist explanation of the practice of intentional psychology because it treats
intentional activity as the outcome of the mind's first order mental states or second order deliberations. Furthermore it claims that this conception of intentional psychology is innate.

I argued in chapters four and five that, while it is certainly true that the child reads meaning into behaviour by applying concepts of mental states, that alone cannot establish the truth of the ToMM claim. ToMM is a specific hypothesis about the nature of the representations on which the ability to read minds depends, a hypothesis which is underdetermined by the evidence. I proposed instead the alternative hypothesis that the child learns to integrate a range of capacities in the project of context construction. The parallel representation of the information on which these capacities depend requires the functioning of somatic markers, representations of body state experienced as primary or secondary emotions. It is the functioning of somatic markers which ensures that relevant information is displayed in working memory in the process of social cognition. I argued that autistic people fail to understand other minds, not because they lack a modular capacity for the formation of theories about the social domain, but because their communicative and interpretative abilities are impaired due to deviant or absent somatic marking. As a result, patterns of behaviour which are the basic data of intentional interpretation are never salient to them as patterns.

The idea is that children can interpret and explain others' behaviour without explicitly representing, consciously or unconsciously, a set of propositions which comprise an integrated theory of human psychology. It is true to say of children that they acquire knowledge of the truths of intentional psychology but that knowledge is practical rather than intellectual because the knowledge is embodied in a set of integrated capacities rather than a single capacity for representation of a set of principles which unify and explain behaviour. In other words, children do not need to explicitly represent a theory of human psychology in order to be good predictors, interpreters and explainers of other people.
The same is true of social cognition in adults. People form beliefs, including second order ones, in order to interpret and explain behaviour in social contexts without explicitly representing a theory of the social domain. I have argued that they do so in much the same way as children learn to negotiate their less complex social environment, by integrating a range of capacities, sustained by a melange of mental representations, in the project of context construction.

The notion of practical knowledge is not behaviourist because it does not reduce knowledge of propositions to a set of dispositions. Rather it treats intentional interpretation as resting on abilities to recognise and respond to patterns in behaviour. ToMM is a specific hypothesis about the nature of that ability which avoids behaviourism by postulating an internal mechanism for theory formation. I have given an alternative account of the nature of the representations on which those abilities depend. However, the rejection of the ToMM model in favour of practical knowledge is not a retreat to behaviourism, rather a demonstration of the virtues of remaining neutral about the nature of actual mechanisms on which practical knowledge depends.
BIBLIOGRAPHY


Currie, G. (Forthcoming). 'Simulation theory, theory theory and the evidence from autism.'

Damasio, A., Damasio, H. & Tranel, D. (1990). 'Individuals with sociopathic behaviour caused by frontal damage fail to respond autonomically to social stimuli.' *Behavioural Brain Research* 41: 81-94.


Dennett, D. (1993) 'The message is: there is no medium' *Philosophy and Phenomenological Research* **53**: 919-931


Fodor, J. (1968). 'The Appeal to Tacit Knowledge in Psychological Explanation.' *The Journal of Philosophy* 65:


Haugland, J. (forthcoming). 'Understanding Dennett and Searle.'


Nichols, S., Stitch, S., Leslie, A. & Klein, D. (Forthcoming). 'Varieties of off line simulation.' :


