PSYCHOLOGICAL REALISM AND THE SIMULATION THEORY OF
BELIEF ATTRIBUTION

by

LISA ANN WOUDZIA

B.A., The University of British Columbia, 1983
M.A., The University of British Columbia, 1986

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Department of Philosophy)

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

April 1997

©Lisa Ann Woudzia, 1997
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Philosophy

The University of British Columbia
Vancouver, Canada

Date April 28/97
Abstract

The presumption of individualism within the philosophy of psychology has been challenged by Tyler Burge who, using arguments like those used by Hilary Putnam against internalistic theories of meaning, points out that the determination of content of one’s beliefs depends, at least in part, upon external factors. With the demise of individualism comes the demise of supervenience, and, one would think, with the demise of supervenience comes the demise of psychological realism. This thesis represents an approach to defending psychological realism in light of the falsity of supervenience.

Psychological realism can be maintained, it is argued, through the adoption of the ‘simulation theory’ which claims that ascriptions of psychological states are based upon a kind of process rather than a theory. This represents an alternative to the traditional view, the ‘theory theory’, which claims that folk psychology is a theory and psychological states are posits in that theory. It is argued that the simulation theory represents a plausible alternative to the traditional view and that this alternative can accommodate the externalist concerns raised by Burge and others without sacrificing realism.
# TABLE OF CONTENTS

Abstract  ii

Table of Contents  iii

Acknowledgement  iv

Chapter One  
Introduction  1

Chapter Two  
The Externalist Challenge to Realism  37

Chapter Three  
Realism, Instrumentalism, and the Theory of Folk Psychology  78

Chapter Four  
Simulation and the Context-Bound Nature of Belief Attribution  116

Chapter Five  
Knowledge, Theories, Information, and Processes  154

Chapter Six  
Autism and the Simulation Theory  184

Reference List  211
Acknowledgement

I am grateful to my Supervisor, Kate Talmage, for her direction, critical comments, and matter of fact attitude about my ability to complete this dissertation. Thank you also to Tom Patton for his insightful and diplomatic comments and for his promptness in reading and returning my work, and to Earl Winkler for his support and for agreeing to be a committee member despite his extremely busy schedule at this time.

Although I am completing my degree at the University of British Columbia, the first half of my Ph.D. student career was spent at the University of Western Ontario. The members of that philosophy department were all very encouraging. I am grateful especially to Ausonio Marras who continued to offer both encouragement and advice on philosophical matters after I had left.

I also wish to thank my husband, Chris Lundquist, for his unwavering faith in my abilities and my son Daniel for simply being himself and giving me perspective. Thank you to my parents, Angelica and Percy Woudzia, who have always been generous in both their emotional and financial support of my education and to John Woudzia for his encouragement. Finally, thank you to my sister, Cathy Nidoski, and my friends, Leslie Holder and Kim Tremblay, for their friendship and interest in my pursuit.
Chapter One

Introduction

This thesis is about the ontological status of intentional psychological states like beliefs and desires that most of us take to be the causes of our behaviour. Do beliefs and desires exist? If so, in what sense? Do they cause behaviour? Most philosophers, and perhaps an even larger majority of non-philosophers, would agree that beliefs and desires exist, that they do cause behaviour, and that they are physically based in the brain or central nervous system. That is, most people are psychological realists. This view has been prevalent and has persisted for the last fifty years or so. Due, however, to challenges that have been put forth by philosophers who have drawn attention to the social, scientific or environmental factors that are relevant to our attributions of psychological states, this view is now in need of either defense or revision.

I will be concerned to defend this view insofar as I will defend the existence of psychological states. That is, I will defend the existence of intentional states as causes of behaviour. I will not, however, defend the mind/brain relationship that has been assumed. Thus, I will argue that the assumed mind/brain relationship is not necessary for realism. In order to address these issues, a number of philosophically technical terms will be put to use. This introduction will be devoted primarily to
explaining these technical terms and addressing matters that are preliminary to the central arguments of the thesis.

Firstly, what is this mind/brain relationship that most people assume is correct and to which I am referring? It has been assumed by many philosophers that if one is to be a physicalist, i.e., not a dualist, then one must be committed, at the least, to the token identity of mental states with physical states. The problem with dualism of course is that it posits 'ghostly' mind or soul states. Since these states are by definition not physical, this type of analysis of mental states cannot explain how mental states could interact with and cause physical states. Since these ghostly states are not acceptable in our scientifically based ontology and since we are confident that mental states have causal force, we have been motivated to find a physically based understanding of the nature of mental states.

This motivation is longstanding. It is demonstrated by the various philosophical positions proposed over the last fifty years that gained wide acceptance in their time because of their apparent ability to analyse the mind in physically based terms. Thus behaviourism, for example, was a compelling view because it exorcised the 'ghost in the machine', as Gilbert Ryle put it. Once it was recognized that behaviourism was fraught with problems an alternative was proposed. The type identity theory, advocated by philosophers such as J.J.C. Smart, U.T. Place, and
David Armstrong, also looked promising, especially if the analysis was limited to sensations and did not include intentional states. With the analysis offered by functionalism came the view that psychology is an autonomous science and, as such, cannot be reduced to anything else. For something to be the belief that P, according to functionalism, that thing must fulfil the causal role of the belief that P and whatever played the causal role of the belief that P was, in fact, the belief that P. Since, however, a variety of things could fulfil the role, the belief that P cannot be identified with any type of physical state. One brain state will play the role of the belief that P in me today and another brain state may play the same role tomorrow, or state X will play the role in me now and state Y will play the role of the belief that P in you now. Regardless of whether functionalism, on any version, is the best way to conceive of the nature of mental states, functionalism enabled philosophers to embrace a new version of physicalism: token physicalism.

Token physicalism is the thesis that although we cannot reduce talk of mental or psychological states to talk of physical states, each instance of a mental state is in fact identical with some token physical state. It is, in hindsight, an amazingly simple and obvious idea. Just as is the case with a variety of other types of things with which we interact in the world, the fact that we cannot reduce the concepts of one type
of thing to the concepts of physics does not mean that the stuff referred to by the concepts are not in fact identical with token states describable at the level of physics. My car, for example, is a Honda Civic that gets great mileage and has good pickup. There is no doubt that my car is a physical thing and that there is some entity that is describable at the level of physics which is in fact identical with my car. But that description would not tell us that the car is a Honda Civic with great mileage and good pickup. It wouldn’t even tell us that it was a Honda. In the same way, the fact that we can’t find any types at the physical level that mirror the conceptual framework of the mental, does not imply that each mental state or entity is not identical with some physical state or entity. My belief that I chose wisely when I bought my car is identical with some physical state of me (the usual supposition is that it is a brain state), but the description of that state in physical terms would not tell us that the state was one of believing that I chose wisely when I bought my car. It would likely not even tell us that it was a state of believing.

The token identity theory is thus a much weaker proposal than either behaviourism or the type identity theory. Actually, the token identity theory is hardly a theory at all. It is more like a statement of what, at least, must be true if physicalism is true. Since it doesn’t propose any reductions one can’t investigate whether the conceptual frameworks match
appropriately. Other than a belief in physicalism, there is no argument for the truth of token physicalism. The truth of functionalism doesn’t even guarantee the truth of token physicalism since, according to functionalism, there is nothing ruling out the identification of the state playing the role of the belief that P with some non-physical state. For the functionalist, it is the role rather than the matter that defines the state: "In functional specification terms, there might be a creature in which pain is a functionally specified soul state. So functionalism opens up the possibility that even if our pains are physical, other pains might not be" (Block 1980, 181).

Token physicalism, then, is the view that every mental event is in fact identical with some token physical event. And this, it has been supposed, amounts to saying that physicalism must be true. After all, as noted above, if it’s not true we end up with ‘ghostly’ soul entities and with no explanation of the causal interaction between the mind and body.

Since it is the psychological level that we are concerned to identify with tokens at the physical level, the token identity relationship is one of supervenience, where the mental states are said to supervene on physical states. For current purposes of discussion we will assume that the physical states we are concerned with here are brain states. The physical level, the
brain state level, is the base level. This means, in part, that there can exist brain states that are not mental states but there cannot exist mental states that are not brain states. Mental states are, thus, dependent upon physical states. Jerry Fodor puts it this way: "States of type X supervene on states of type Y iff there is no difference among X states without a corresponding difference among Y states. So, in particular, the psychological states of organisms supervene on their brain states iff their brains differ whenever their minds differ" (1987, 30). Thus, there can be no change in mental state without some change in brain state. Supervenience asserts no more about the physical nature of mental states than the token physicalist identity theory and is thus not a stronger thesis. It does, however, describe the nature of the relationship between the mental and physical better than the token identity theory because it captures the notion of dependence. This in fact was Jaegwon Kim's aim (1982) when he first applied the notion of supervenience to formulate the thesis of psychophysical supervenience. Thus, since supervenience asserts nothing stronger than token identity, those who profess a commitment to token identity would also be committed to supervenience.

To assert that supervenience is false is to deny the truth of token physicalism and, if one assumes that dualism is false, it is also to deny the existence of mental states. If one assumes that dualism is not an option then in order to maintain that
mental states exist they must be physically realized. If mental states are not realized physically, that is, if they do not supervene upon the physical, then, assuming dualism is false, they cannot exist. We shall henceforth assume that dualism is false. Realism is the position that mental states do exist. Anti-realism is the position, of course, that they don’t exist. Whether one’s analysis of the nature of mental states commits one to being a realist or an anti-realist, however, is much less straightforward. Before we consider the less straightforward cases of realists and anti-realists, let’s consider a simple case of anti-realism.

Eliminative materialism is the clearest example we have within the philosophy of mind of anti-realism about the mental. Eliminativists view the concepts we use in our discussions of the mind as analogous to concepts that we have, over the course of advancement in science, decided to abandon. We abandoned them because we found that they didn’t actually refer to anything. Examples cited by Paul Churchland (1984, 43-45) for example, include the concepts of witches and phlogiston. We did not simply revise our concepts in these cases. Since the whole framework within which these concepts were embedded was mistaken they were beyond revision. They had to be abandoned. What we now view as instances of psychoses of various kinds we once viewed as demonic possession:
That witches exist was not a matter of any controversy. One would occasionally see them in any city or hamlet, engaged in incoherent, paranoid, or even murderous behaviour. But observable or not, we eventually decided that witches do not exist. We concluded that the concept of a witch is an element in a conceptual framework that misrepresents so badly the phenomena to which it was standardly applied that literal application of the notion should be permanently withdrawn. The concepts of folk psychology—belief, desire, fear, sensation, pain, joy, and so on—await a similar fate, according to the view at issue. (Churchland, 1984, 44)

Churchland would deny the thesis of psychophysical supervenience because, according to eliminative materialism, there are no such things as beliefs and desires, so they can't supervene on anything. Thus, the thesis of psychophysical supervenience is false. Notice that this is different from saying that beliefs and desires don't exist because psychophysical supervenience is false. Churchland does not argue to the eliminativist conclusion based on the denial of the truth of supervenience. He argues on the basis of the failure of type identity theory, on the basis of what he sees as historical parallels like witches and phlogiston, and on the basis of what he sees as the explanatory poverty of folk psychology. Thus, one could reject eliminativism by rejecting Churchland's views on these matters. One could deny the significance of the failure of type-type identifications, one could reject his choice of historical parallels, and one could argue that progress has in fact been made and that our folk psychological framework is actually a pretty good theory for explaining and predicting behaviour. In other words, the arguments for eliminativism seem to be a matter of perspective.
Some people find them persuasive, others do not. One is not compelled by logic to accept them. Among those concerned to defend realism, then, the arguments standardly presented by those defending eliminative materialism have not generally caused much concern.

If eliminative materialism on its own does not compel us to abandon realism, what is the problem? It seemed that, although we couldn’t naturalize the mental realm through behaviourism or the type identity theory, supervenience and token identity do allow us to hold that the mental is physically realized. But there is a problem. Even this weakest form of physicalism has been challenged. It has been challenged by the claims of a view called ‘externalism’. Externalists argue that at least some of the determinants of meaning in language are external to the individual using the language. And the corollary in the philosophy of mind, what Tyler Burge (1979) calls ‘anti-individualism’, is that the contents of one’s propositional attitudes are determined at least in part by external factors.\(^1\)

Our purpose will be to show that, regardless of what type of external factors play a role in determining content, psychological realism is still a viable position. That is,

\(^1\) 'Externalism' is commonly used to name the position at issue within the philosophy of language. Burge is more specifically concerned with psychology and with the problems with the theory of individualism with psychology. I don’t, however, see any relevant reason for not using the terms interchangeably in our discussion here.
realism can be maintained despite the fact that not all the contents of my mental states supervene on my brain or bodily states. Before we consider how this might be done, let's review the externalist arguments.

Two of the most influential brands of externalism are those defended by Hilary Putnam (1975) and Tyler Burge (1979). Burge's externalism is focused upon social facts. Putnam's externalism is based upon scientific facts about the objects of the world. The story most often retold in support of an externalist view is Putnam's now very familiar Twin Earth thought experiment. In his thought experiment Putnam asks us to consider a planet that is exactly like Earth in every respect but one. The liquid called 'water' on Twin Earth has a different chemical composition than does the liquid called 'water' on Earth. To abbreviate, we shall say 'water' on Twin Earth is XYZ, while water on Earth is H₂O. The chemical composition is the only respect, however, in which water and water(te) differ. People on Twin Earth drink XYZ, their lakes and oceans are full of it, they sprinkle their lawns with it, etc.

Next we are to consider an inhabitant of Earth, Oscar, who has an exact replica of himself, a doppleganger, living on Twin Earth. Everything Oscar₁ (Earth Oscar) does and says, so too does Oscar₂ (Twin Earth Oscar). Being exact replicas, Oscar₁ and Oscar₂ have backgrounds or histories including relationships
with people and the world, accomplishments, interests, etc., that look qualitatively exactly the same and neither one of them has any knowledge of chemistry. Now when Oscar1 says, "I sure could go for a glass of cold water", he is uttering a word with a different extension and, hence, different meaning from Oscar2 who says, "I sure could go for a glass of cold water." Given that Oscar1 and Oscar2 are exact replicas of one another, there is nothing that is internal to them that could account for the difference in meaning. The word 'water' means something different when uttered by Oscar1 than when uttered by Oscar2, because the word has a different extension on their respective planets. This is true, Putnam claims, regardless of not only the Oscars' knowledge but of everyone's knowledge of whether the chemical compositions of water on the two planets differ. Thus, in 1750, before the chemical composition of water was known on either planet, Oscar1 and Oscar2 would have still meant different things when they said, "I sure could go for a cold glass of water", because even though no one knew it, the extension was just as different then as it was after we discovered the difference in chemical composition.

Burge offers a different brand of externalism and he has his own stories to tell which don't depend upon natural kind terms like 'water'. Burge's most familiar story involves the concept of 'arthritis'. In the example the subject seems to have a good understanding of the meaning of 'arthritis' and is suffering
pain in his thigh. The person in question, we’ll call him Daniel, correctly believes many things about arthritis. Daniel believes, for example, that he has suffered from arthritis for years, that the arthritis in his wrists and fingers is worse than the arthritis in his ankles, that it is better to have arthritis than cancer of the liver, etc. He can carry on perfectly rational and intelligent conversations with people regarding arthritis. In addition, he thinks that he has arthritis in his thigh.

Daniel visits the doctor and complains that his arthritis is getting worse and that he is now experiencing arthritis in his thigh. The doctor explains to him that arthritis is an inflammation of the joints and that it is not possible to have arthritis in the thigh. Daniel accepts the doctor’s word and corrects his complaint to say that he is experiencing pain in his thigh.

In the counterfactual situation that Burge asks us to consider, Daniel’s history is exactly the same as explained above up until the time he visits the physician. He has the same beliefs, he learned about arthritis in the same way, he carries on the same conversations regarding arthritis. At the moment he says to the doctor that his arthritis is getting worse and that he is now experiencing arthritis in his thigh, there is no difference between Daniel-a (actual world) and Daniel-cf (counterfactual
world). We are to imagine, however, that in this counterfactual world, 'arthritis' does not refer to an inflammation of the joints, but refers to a variety of rheumatoid ailments. The doctor thus does not correct his claim regarding arthritis and in fact begins to discuss treatment.

In the actual case Daniel has made an error. In the counterfactual case he has not. Burge's conclusion from this story is that the content of the utterance or thought, "I have arthritis in my thigh", is different for Daniel-a and Daniel-cf and the difference is not reflected in their internal physical or individual historical states but is a consequence of the external, social environment. 'Arthritis' in the counterfactual situation doesn't mean arthritis, but means something else entirely, like 'tharthritis', which refers to a general rheumatoid ailment. The conclusion, as with Putnam's Twin Earth thought experiment, is that meaning is determined by factors external to the individual, except that in Burge's example the externalism runs right through to thoughts whose content is determined in part by social factors. What 'arthritis' means appears to depend upon what we conventionally take it to mean rather than upon some underlying natural kind. Conventionally, we happen to take it to mean what experts, in this case medical professionals, tell us it means.
Burge points out that the argument has wide application and does not depend on using an example of the same type as ‘arthritis’, nor does it depend on experts. He makes the following claim: "The argument can get under way in any case where it is intuitively possible to attribute a mental state or event whose content involves a notion that the subject incompletely understands" (1979, 79). He cites various examples including sofas, beef brisket, clavichords, and the legal notion of a contract.

Putnam and Burge both offer what seem to me to be convincing examples of how the determination of meaning must be influenced by factors outside of the individual. Although these are clearly different brands of externalism, the important issue is whether meaning or propositional attitude content is determined by anything outside of the individual. Externalism threatens realism because of its denial of mind/brain supervenience. If the contents of beliefs can vary because of external factors, while the internal factors, the states of an individual’s brain, remain the same then the contents of beliefs do not supervene on brain states. If externalism is right, then supervenience and token physicalism are false and we are left once again with the mind/body problem. It seems that we now need a new explanation of how mental events can indeed be physical events. But if supervenience is the weakest form of the identity theory, then it is difficult to see how we can remain optimistic about this
project. Unless we can come up with some alternative idea it seems that we have no choice but to deny realism about the mental.

Although some philosophers, notably Stephen Stich (1996, 168-91), don’t think that naturalizing intentionality, that is, finding a physicalist understanding of mental states like beliefs and desires, is necessary to justify realism about those mental states, most still do. As a testimony to the strength with which this conviction is held, Stich quotes Fodor:

> if it isn’t literally true that my wanting is causally responsible for my reaching, and my itching is causally responsible for my scratching, and my believing is causally responsible for my saying...if none of that is literally true, then practically anything I believe about anything is false and it’s the end of the world. (1990, 156)

Of course Fodor is being characteristically overdramatic here, but he is still serious about his point. This being the case Fodor and others have attempted to come up with a solution to the problem posed by externalism.

A proposal that is consistent with supervenience and an individualist view of psychology is to separate the content of mental states into two types: broad and narrow. Broad content

---

includes external factors whereas narrow content includes only what is internal to the individual. Thus, although it is the case that supervenience is violated in the case of broad contents, it is not violated for narrow contents. And, it is further claimed, it is narrow contents that are important to psychology since those are the contents on the basis of which we behave.

Assume, for example, that I don’t know that Jane Austen is the author of *Emma* but I have heard that *Emma* is an excellent book. Furthermore, I know that Jane Austen wrote *Sense and Sensibility* and I thoroughly enjoyed reading it. I had decided that I would like to read another of Austen’s books but that I would first read something from the author of *Emma*. I enter a bookstore and ask the clerk if she has any books by the author of *Emma*. I am handed a copy of *Pride and Prejudice* and when I check to see who wrote it I am surprised and somewhat confused since I didn’t tell the clerk that I wanted something by Austen. I repeat my request for something by the author of *Emma* and the clerk then informs me that Jane Austen is the author of *Emma*.

My reaction to the clerk and my behaviour of repeating my request was dependent upon my understanding of 'Austen', which did not include the fact that she is the author of *Emma*. The cause of my behaviour of asking for a book by the author of *Emma* was my desire for a book by the author of *Emma*. Broadly
construed, of course, the author of *Emma* is Austen, but narrowly construed, that is, construed only in terms of what I know, the author of *Emma* is not Austen. When attempting to explain my behaviour it will, it is claimed, be the narrowly construed content that we should be interested in. If we construe my state of mind in broad terms it will not explain my behaviour. If we construe my state of mind broadly, it is unclear why I would have been surprised and why I would have repeated my request. For the purposes of psychology then, it seems that only narrow content, the content based upon what is going on in the individual, is relevant.

This analysis, which depends on Frege's sense/reference distinction, certainly has some intuitive appeal. We cannot deny that there is a difference between sense and reference. But it seems to me that this, in itself, is not a reason to suppose that there are such things as narrow contents upon whose basis we behave. The explanation of my behaviour in the above example will not, it seems to me, depend on anything like narrow content. That my behaviour is explained by the fact that I didn't realize Austen was the author of *Emma* means that I missed something significant about the broad meaning or reference of 'Austen'. This does not mean that my behaviour is thus best explained by using narrow contents. It simply means, it seems to me, that my behaviour is best explained in broad meaning terms with the added piece of information that I didn't realize that
Austen was the author of *Emma*. My behaviour is explained in terms of something like making a mistake. But because I am mistaken or ignorant of some facts about Austen, does not mean that I do not think of Austen in a meaningful way. In a way, that is, that is meaningful according to broad criteria. In fact, as I explain below, the problem with the notion of narrow content is that narrow content does not allow the individual to act on the basis of the contents of his or her beliefs when those contents are understood as bearing meaning. Since, however, the idea of narrow contents has initial intuitive appeal and it would enable us to maintain supervenience, it is worth investigating further.

The difference between broad and narrow content is sometimes discussed in terms of a difference between a theory of meaning and a theory of psychological content. It is claimed that while broad content is relevant to determining meaning it is not relevant to determining the contents of belief states. This distinction, it seems to me, reveals the problem with narrow content. It appears to offer a straightforward reason why narrow content cannot be a defensible approach to defending psychological realism. The question is: if the inclusion of broad content is relevant to meaning then how can an explanation of behaviour be meaningful if it cannot include such content? How can ascriptions of propositional attitudes ever be correct if the meaning of those ascriptions depend upon broad criteria,
but the contents do not? And if our ascriptions cannot be correct then in what sense are we able to maintain realism about psychology?

If my behaviour in the bookstore is explained by saying that I believed, when I asked for a copy of a book by the author of Emma, that I was asking for a book by an author other than Austen, it is unclear, on the narrow psychology account, how to capture the meaning of this content. If we cannot look to factors outside of the individual to determine the contents of the individual’s beliefs, then it is not clear how those contents can be meaningful at all. Although I may not have known that I was referring to Austen when I asked for the book, I was intending to refer to someone about whom the only thing I know is that she (or he) wrote Emma. Thus, it seems to me, I was intending to refer to someone about whom there are many true statements of which I am unaware. For example, even though I didn’t know it, it seems to me that I was referring to someone who was born in 1775, was the daughter of a clergyman and was also the author of Mansfield Park. If the contents of my beliefs are determined narrowly, however, since I am unaware of any of these facts about the author of Emma, they cannot be part of the content of my belief. But if the content of my belief is not partly determined by any facts of which I am unaware, then, given that I am unaware of any facts about the author of Emma, how does the content of my belief have any meaning at all? The
problem, then, is firstly how we separate the narrow contents from the broad and, secondly, once we have done so, whether we are left with realism about beliefs as we currently understand them or realism about some other notion of beliefs.

It is generally agreed that our ordinary intentional psychological concepts are not individual or narrow as they now stand. Thus, as they now stand, they are inappropriate for the purposes of explaining behaviour. The individualist proposal then, is not a proposal about how to analyse current folk psychology, but is a claim about the way it should be. It is a revisionist proposal. Robert Stalnaker describes the move of supporting narrow contents in response to the externalist points in this way:

The revisionist response makes a negative and a positive claim. The negative claim is that no systematic explanatory theory of behavior will be tenable unless it is individualistic. The positive claim is that although ordinary intentional psychological concepts are not individualistic as they stand, they can be revised in a way that renders them individualistic while preserving the basic structure of intentional explanation. (1989, 289)

Stalnaker then discusses who is defending and rejecting these claims. Jerry Fodor, who endorses realism, the position I am concerned to defend in this thesis, defends both the negative and the positive claims. Daniel Dennett, who is considered an instrumentalist rather than a realist, defends the positive claim. And Stephen Stich (who has recently changed his mind) and
the Churchlands have defended the negative claim while rejecting
the positive claim which is consistent with their anti-realist,
eliminative materialist position. (We have been taking it for
granted that realism requires the truth of supervenience which
is just what the negative claim presupposes).

Stalnaker argues that it isn't as easy as it is often made out
to "factor out the 'organismic contribution' to an intentional
mental state" (p.289). I agree. In fact, I reject both the
negative and the positive claims. Since I reject the negative
claim that individualism must be true for psychology to be a
viable science of the explanation of behaviour, I do not need to
defend a particular version of narrow content in order to defend
realism. Just how it is that psychological realism can be
compatible with the denial of individualism is the issue that I
must address and that is, in fact, the central issue of this
thesis.

There are, then, two routes to the defense of psychological
realism in response to the externalist claims. One, that we have
been discussing above, is to claim that psychology is interested
in narrow contents only and that current psychology can be
revised to accommodate this fact. The second option, the one I
shall pursue, is to claim that psychology is, as our everyday
ascriptions assume, concerned with mental states, broadly
construed. Understanding how psychology could be concerned with
broadly construed states and still be defensible as a science of behaviour is not a simple task. It will involve a reconception of the role of the external environment in our ascriptions.

Philip Pettit and John McDowell in their "Introduction" to "Subject, Thought and Context" (1986) put the issue in terms of a radical and a less radical option. They see the position I will be defending, the position that psychological ascription is and should be broad, as the more radical option. The distinction between the radical and the less radical views is found in their respective responses to the mutually accepted claim that context partly determines (broad) content. The less radical response, they say, accommodates the claim that context partly determines content by:

regarding the environment, social and physical, as straightforwardly external to a realm of context-free psychological facts. In the more radical response, in contrast, the notion of context undergoes a shift. We can no longer regard the social and physical environment as simply surrounding the psychological subject. Rather, we have to accept that contextual facts inextricably permeate the field of psychological investigation, even when what is under study is the psychological organization of an individual. (1986, 14)

Pettit and McDowell also offer the following as a potential objection to the less radical version which uses narrow contents to protect a conception of individualistic psychology as theorizing about context-independent mental states:
...if we leave the communal environment,...out of consideration,...then our picture will contain nothing at all that is recognizable as the subject of mental states....there will be no obvious reason in that case to accept that the findings of such a discipline would have any constitutive relevance to the mind. From this standpoint, the postulation of 'narrow content'...will look like a self-deceptive attempt to conceal the disappearance of the cognitive subject from the picture. (1986, 14)

This objection rests on the view that Pettit and McDowell discuss that to mean something by a word, one must regard oneself as responsible to a norm which can only be understood in a communal context.\(^3\) It is because I endorse this sort of view that I reject the narrow contents route to solving the problem posed by externalism. If it is agreed that meaning depends, in part, upon external factors, then by bracketing out those factors we are bracketing out at least part of the meaning. And the part of meaning that is being bracketed out is the part that is required for understanding—the part, that is, that is essential to communication because it is shared and it is shared because it is determined externally, not individualistically. The problem is that, once we do this bracketing out, it isn’t clear what we have left. It isn’t clear what narrow content is.

It seems to me that if we exclude broad content we thereby exclude mental or psychological states. And since the question is one of realism about psychological states this approach will

---

\(^3\) They see this view as offered by one reading of Wittgenstein’s reflections on rule following and his argument against the possibility of a private language (p.10).
not give us the results we want. It’s true that we are interested in analysing psychological states understood as causal states, i.e., those states that cause our behaviour. But that analysis does not exhaust the nature of these states. We commonly hold mental or psychological states to be meaningful and to be about something. In fact, the quality of being ‘about something’ has been noted, by Brentano, to be a distinguishing mark of the mental. But being about something seems to imply a connection to the external world. Narrow contents are, by definition, determined only by internal states. So if a narrow content is about anything, it must be about an internal state, not a state of the world. It is clear that on the broad reading, my belief that ‘water is wet’ is about water. It is not clear, on the narrow reading, what it could be about. It seems to me that once we exclude broad content, the states can’t be about anything.

The attempts to define narrow contents don’t generally follow the pattern of reasoning that I have just pursued. Nonetheless, I believe that the arguments they offer fail for the reason I give, i.e., that it is not clear what narrow content could be about and, if this is unclear, psychology cannot and should not be based upon narrow content. In defense of this claim I will discuss Stalnaker’s comments on two accounts of narrow contents that have been proposed: those of Jerry Fodor and Daniel Dennett. Fodor proposes that narrow content is to be understood
in terms of a relativizing of the connection between content and truth conditions to context. Dennett introduces the idea of 'notional worlds' wherein my doppleganger and I would be believing the same thing when we believed that 'water is wet'. With this Dennett is not, like Fodor, changing our notion of the content of a belief itself but, rather, is proposing an alternative relation between a believer and a context in virtue of which the content correctly describes the believer's beliefs. Neither of these accounts, it seems to me, is satisfactory.

Stalnaker argues, in effect, that Fodor's account is empty. Fodor claims that the Twin Earth cases don't break the connection between content and extension, they just relativise that connection to content. He argues for the following 'extensional identity criterion' for narrow content:

...there's something about the relation between Twin-Earth and Twin-Me in virtue of which his 'water'-thoughts are about XYZ even though my water thoughts are not. Call this condition that's satisfied by (Twin-Me Twin-Earth) condition C. Similarly, there must be something about the relation between me and Earth in virtue of which my water-thoughts are about H2O even though my Twin's 'water'-thoughts are not. Call this condition C'. Short of a miracle, it must be true that if an organism shares the neurophysical constitution of my Twin and satisfies C, it follows that its thoughts and my Twin's thoughts share their truth conditions. But now we have an extensional identity criterion for mental contents: Two thought contents are identical only if they effect the same mapping of thoughts and contexts onto truth conditions. (Fodor, 1987, 48)
Stalnaker rightly points out that Fodor does not tell us how a mapping from context onto truth conditions can be specified. He also argues that Fodor's argument "obscures the fact that it is a substantive hypothesis that the internal states of believers contain thoughts that determine such mappings" (p.295). Fodor doesn't really offer an argument. He simply points out that if the external world were the same and the neurophysical state of the believer were the same, the content would be the same: "That is only to say that truth conditions are determined by the conditions that are relevant to determining them" (p.295). Stalnaker offers a parody of Fodor's narrow content criterion. He asks us to suppose that he has the property of being three miles from a burning barn. We are to suppose further that there is a counterfactual situation in which he has a counterpart who is located at the same place except that there is no barn and thus his counterpart doesn't have this property. His counterpart is instead three miles from a snow-covered chicken coop.

Now there is presumably something about the relation between my counterpart and his world in virtue of which he is three miles from a snow-covered chicken coop even though I am not. Call this condition C. Similarly, there is something about the relation between me and my world in virtue of which I am three miles from a burning barn, even though my counterpart is not. Call it C'....Short of a miracle, it must be true that anyone in the location that both I and my counterpart are in in our respective worlds would be three miles from a snow-covered chicken coop if condition C obtained and three miles from a burning barn if condition C' obtained. (1989, 295-296)
The problem here is that every location is such that if certain external conditions obtain, then that location is three miles from a burning barn. There is no reason to think that the intrinsic location should be identified with some function that determines the relational property of being three miles from a burning barn. And there is no reason to believe that some intrinsic property of the individual can be identified with some function that will determine how thoughts are to be mapped onto contexts, and will thus determine the broad, relational content.

Fodor admits that narrow content is not something that can be expressed. We can only 'sneak up on it' (1987, 51). The only way we can talk about content is to express it in a sentence of an Earth-based language. And the meaning of a sentence in English, for example, is partly determined by external factors. But, he does claim that expressions in English can be used to pick out narrow contents "via their hypothetical semantic properties....the narrow thought that water is wet is the unique narrow thought that yields the truth condition H₂O is wet when anchored to my context and the truth condition XYZ is wet when anchored to his" (1987, 51). This, however, leaves us with the puzzle of how to determine the anchoring conditions. It seems to me that this is just another way of saying that condition C or C' obtains. To say that content is anchored to context is not to give an argument in defense of narrow content. After all, that content is partly dependent on context was the whole point of
the Putnam/Burge stories. That there is some narrow content that gets anchored and then becomes the broad content is a substantive hypothesis for which Fodor has not, it seems to me, given explicit arguments.

Unless narrow contents contain within them some specification of how they are to map contexts onto truth conditions it is not clear how such contents can have any meaning. And proposing that there is some condition C such that, if it obtains a certain mapping will obtain, does not illuminate how the mapping will obtain. Furthermore, the whole idea that this mapping function could be contained in narrow contents is difficult to grasp. Putnam, for example, claims that he doesn’t understand, given Fodor’s view that narrow content is a function from context to referent, how such a function could be in the speaker’s head (1988, 43).

'Ausonio Marras (1989) defends Fodor’s position and claims that Fodor’s view is defensible on the basis of the fact that he relies on his representational theory of mind (RTM). It is the internal sentences, which are considered beliefs because of their internal functional role, that are the things that get anchored in a context. It seems to me, however, that this will not answer the problem of how narrow contents, even if they are internal sentences, get anchored.

Marras also supplements Fodor’s account by claiming that it might be possible that narrow contents can be expressed by ordinary that-clauses by using meta-linguistic descriptions. It seems to me, however, that meta-linguistic descriptions can only be used in those cases like Twin Earth where we already know that supervenience has failed for the regular ascriptions and we know what sort of thing is held in common between Earth and Twin Earth—in this case the phenomenological properties of the stuff inhabitants of both planets refer to as ‘water’. Without this information, we wouldn’t know what the meta-linguistic ascription
As we shall see in chapter two, Fodor has, subsequent to his above analysis of narrow contents, approached the defense of psychological realism from another angle. He doesn't claim to have changed his mind about narrow content but he does claim that narrow content is superfluous and therefore unnecessary to a defense of realism. While I will disagree, once again, with Fodor's attempts to defend realism, I do agree with his move away from the emphasis on narrow contents. This is because it seems to me that one can either talk about brain states or contents. If we are interested in doing psychology we will be talking about contents. Once we start talking about contents we are talking about meaningful states and such states are partly determined externally. If we remove the external contribution, we are left, it seems to me, not with narrow contents, but with brain states. And as Pettit and McDowell point out, without meaning, we have no subject. Without a subject, interacting with a world and with others, we have no psychology. 5

should be about. And giving meta-linguistic analyses of every term used in our ascription would result in a regress in which any meaningful term gets analyzed out. For details on this argument see Loar, 1988, p.130.

5 Of course, had Fodor's suggestion worked, this would not have been the case. Since for Fodor, narrow contents would be 'anchored' and we could not escape context, our ascriptions would be about narrow contents, but would be expressed in terms of broad contents. Meaning and the subject would not be lost on Fodor's account since the link between the person's internal state and the state of the world is part of the analysis. Once again, however, how the individual internal state of a person can contain a function which maps content onto context and truth conditions is a mystery.
Fodor's proposal for narrow content depended on a reconception of narrow content as some kind of function. The function would effect different mappings in different contexts. Dennett's proposal, alternatively, suggests that the narrow content remains the same throughout different contexts. To explain his position he introduces the idea of a notional world in which the semantic properties of our internal states are fixed. These semantically evaluable fixed states are the narrow contents:

"...nowhere is it written that the environment relative to which we fix a system's semantic properties must be a real environment, or the actual environment in which the system has grown up" (1987, 154). Notional worlds are developed by looking at the internal state of an individual and deciding what kind of world to which that individual is best suited. It is like coming upon some antique gadget whose purpose is unclear and trying to determine, based upon its physical non-relational characteristics, for which tasks it is best suited. We imagine a world--a notional world--in which, given the facts apparent to us now, the object functions excellently. Dennett suggests that we apply this exercise to ourselves. In doing so we come up with a solution to the problem of narrow content because my twin and I will end up in the same notional world. Referring to human beings, Dennett says:

Such organisms have internal structure and dispositional traits so rich in information about the environment in which they grew up that we could in principle say: this organism is best suited for an environment in which there is a city called Boston, in
Dennett then says that this is of course a myth, but a theoretically important myth, because it reveals how, if one believed in determinate narrow content, one would have to go about finding it. Dennett himself is considered an instrumentalist. He would say that there may be many notional worlds that would match any particular individual state and that there is no fact of the matter about which one is the right one. So, in notional world A, John’s state will have one content and in notional world B his state will have another content. Still, he does seem to be suggesting that we can determine which worlds the state is best suited for simply by looking internally, that is, without bringing considerations of the external world into the picture. In doing so, we are being neutral about the environment. Stalnaker puts it this way:

Dennett’s project might be seen as an attempt to eliminate the context-dependence by defining content relative to an absolutely neutral context that is free of all presuppositions about the external environment. But on the causal informational account of representation, informational content is essentially relative to a range of alternative possibilities that are determined by general facts about the causal structure of the world in which the organism functions. (p.27)

I agree with Stalnaker that Dennett has a problem. The account of narrow content that Dennett gives depends upon there being
some context-neutral notional world. There may be a notional world that is neutral between Earth and Twin Earth, but that doesn't mean that it's completely context-neutral. The issue can be put in terms of whether your internal state determines the world or the world determines your internal state. If it is the case that what your internal state means depends, in part, upon the external world then the notional world exercise, in which the world that is conceived of is the world that will be best suited to the internal state, is not a possible exercise: "It is internal states of the representor, on this kind of account, that contain information (or misinformation) but the system of causal relationships in virtue of which those internal states contain information cannot itself be something internal to the representor" (Stalnaker 1989, 306). For example, it may be that the belief that water is wet is equally well suited to Earth and Twin Earth and, thus, if we were to construct a notional world based upon the usefulness of that belief we would not distinguish between Earth and Twin Earth. The problem is that we began the example with a belief with content. How, if we are excluding the external world, do we tell what the internal state is in order to determine what world it is best suited for? The internal states contain information, but only in virtue of the causal relationships it has with a particular environment.
There are, of course, other proposals for narrow content. I shall not, however, discuss any others here. I agree with Stalnaker when he concludes his discussion of Dennett with the claim that no reasonable notion of content can result if we ignore information about the believer's historical properties and his relation to his environment (p.306). Since this is what any proposal about narrow content must do I simply do not see how any such proposal can work as the basis for psychology. Psychological realism, it seems to me, must find its defense elsewhere.

The more radical option, as Pettit and McDowell put it, is to defend realism by accepting that contextual facts permeate psychological attributions. If we defend narrow contents we preserve supervenience, but it seems to me that the cost of doing so is abandoning the subject and therefore abandoning psychology. The alternative, however, does not allow us to maintain supervenience. And the cost of abandoning supervenience, it has been argued, is to abandon the causal interaction of beliefs with behaviour thereby abandoning the basis for claiming psychological realism in the first place. In addition, it seems that if you abandon the idea that mental states supervene on physical states you are left with two

6 See, for example, Ned Block, "An Advertisement for a Semantics for Psychology", in Midwest Studies in Philosophy, 10. For a critique on this view, as well as on another of Fodor's versions of narrow content, see Hilary Putnam's Representation and Reality, MIT Press, Cambridge Mass., 1988.
unattractive options: dualism or eliminative materialism. As was discussed earlier, supervenience is the weakest form of an identity theory and, thus, in any physicalist story that includes mental states supervenience must hold. Since the reconciliation of realism with externalism is a complex issue and is the focus of the rest of the thesis I will not deal with it here. The response to the second concern, however, is fairly straightforward.

The manner in which I defined supervenience earlier was somewhat misleading and oversimplified. My discussion was limited to individual supervenience. One can claim, however, that mental properties supervene on physical properties without agreeing to the claim that mental properties supervene on physical properties of the brain or body. Mental states can supervene on physical states outside of the body. One can hold supervenience to be true without specifying how it is true. That is, one need not know what, in particular, mental properties supervene upon to know that they do supervene on physical properties. To hold such a view would simply be to hold that given all the physical properties of the world, if those physical properties were reproduced in some new world, the mental properties of the original world would also be reproduced. There could not be different mental properties in the new world given that all the physical properties are the same.
This response assumes that the charge that to deny supervenience leads to either dualism or eliminative materialism is correct. The response claims that in denying individualism one need not deny supervenience. An alternative response would be to deny that any kind of supervenience is required for realism. Stephen Stich (1996, 186-189) makes this kind of argument. Imagine a world, call it W1, which was created just a few seconds ago but, although it has no history, it is exactly the same physically as our world. He thinks that in such a world intentional properties, because of their historical relationship with the world, cannot supervene on physical properties. Thus Laurence can have thoughts about Julius Caesar because he is connected in the right way to Caesar, whatever that way may be, but in W1 Laurence has no beliefs about Caesar. This, Stich claims, should not worry us: "There are lots of properties that do not globally supervene on physical properties—the property of being a genuine Picasso, for example (there are no Picassos in W1) and the property of being a real $100 bill. But from the fact that these properties do not globally supervene on the physical, it surely does not follow that there are no real Picassos and $100 bills in our world" (1996, 187).

Whether the analogy Stich offers constitutes a good enough argument to dismiss the worry about the incompatibility of supervenience with realism is something I will discuss further in the next chapter. At this point, I simply wanted to note that
there are different kinds of supervenience and that denying supervenience does not, in itself, lead one to a forced situation in which the only choices are dualism or eliminative materialism. The real problem, however, is that of explaining how, if individual supervenience is false, intentional psychological states can cause behaviour. And for realism about psychology, we require no less than such an explanation of causal interaction.
As we saw in Chapter One, on the face of it, it appears that the acceptance of any type of externalist view with regard to content of intentional states will result in a denial of realism about intentional states. Despite this apparent disparity, however, at least three prominent contemporary philosophers, Tyler Burge (1979), Jerry Fodor (1995), and Stephen Stich (1996) want to maintain both. Since it is the project of this thesis to defend the view that both can be maintained, the arguments presented by these philosophers will receive careful attention. Ultimately, however, I don’t think any of their arguments work or, at least, are strong enough to warrant the conclusion that realism is not incompatible with externalism.

Realism asserts the physical reality and causal efficacy on behaviour of intentional states such as beliefs and desires. Standardly this has been interpreted to mean that some form of supervenience must be true. In order for it to be the case that I act on the basis of my beliefs and desires, those beliefs and desires must be physically realized in me. If they are not, then it is not possible for my behaviour to be accounted for in terms of my beliefs and desires. The elementary argument is as follows: The first premise is that realism entails that the contents of intentional states be physically realized within the
individual. Without this we could not legitimately say that one's beliefs and desires cause his or her behaviour. The second premise is the externalist claim that the contents of intentional states are not fully or necessarily physically realized within an individual. Individual supervenience is false. The conclusion then is that since realism entails supervenience and supervenience is false, realism must be false.

Although this would appear to be the position that the externalist must end up in, Burge attempts to defy the conclusion and denies supervenience at the same time he asserts realism. He would reject this argument by denying premise one. Fodor, on the other hand, would accept premise one and would attempt to minimize the force of the argument by qualifying the claim made in premise two. Fodor would accept that there are occasions on which intentional states are not physically realized, but that these occasions are irrelevant or accidental and do not warrant the generalized claim made in premise two. Stich offers a combination of responses that ultimately, it seems to me, depend on the acceptability of narrow contents. We will consider Fodor's arguments first.

In "Life Without Narrow Content?", chapter two of The Elm and the Expert, Fodor attempts to reconcile externalist examples with what he calls computational semantics. Fodor does not claim to have abandoned his view on narrow contents. Recall that the
narrow contents response to the externalist examples is to acknowledge that perhaps the externalist examples count against meaning being individualistic, but they needn't count against intentional states. This, I argued, leaves one with a psychology that looks nothing like we thought it would at the outset. It leaves one with a psychology bereft of meaning. Thus on a narrowly construed psychology, what is important to explanations of behaviour is not the intentional states as we conceive of them in folk psychology. Behaviour will have to be explained in another way that does not involve the ascription of meaningful intentional contents because such contents would be determined at least in part by factors outside of the individual; in other words, they would be considered broad contents rather than narrow. Broad includes external factors, narrow contents are sensitive only to what goes on within the individual. But since Fodor's focus in "The Elm and the Expert" is not upon how narrow contents can give us realism in psychology, we will now consider his alternative view of how to maintain realism and accept broad (externalist) contents.

Fodor's position is that although we may in fact determine contents according to broad criteria, we still don't actually experience the problems that are indicated in the Twin Earth type examples and, thus, externalism poses no serious threat to
realism.7 "What I am going to argue is this: the considerations that have been supposed to show that an externalist construal of content won’t meet the purposes of psychological explanation are, on balance, unconvincing. So maybe narrow content is superfluous" (1994, 28).

Fodor presents three arguments against the twin type cases: 1) The H2O/XYZ story used in Putnam's example is not nomologically possible; 2) There are cases of twins that are nomologically possible, but they are accidents and, as such, do not impugn our theories, and; 3) There are still other apparent cases of twins that are nomologically possible and not accidents. These cases which involve using experts turn out not to be genuine examples of twins because the use of experts invokes a distinction between the so-called twins. Fodor takes these arguments to have shown that we needn’t worry about twin cases because there really aren’t any. Despite Fodor's reassurance, however, I'm still worried.

Fodor's first argument in which he points out the fact that the H2O/XYZ story is not nomologically possible is not news nor, it seems to me, is it an argument. Granted, the story does not show that it is empirically possible for broad contents to vary

7 Actually, Fodor doesn't use the word 'realism' in his discussions and I am interpreting him in this way. What he sees himself as reconciling is the intentionality of psychological laws with the computational/information processing model of mind. See The Elm and The Expert, MIT Press, 1994.
between me and my doppleganger; it simply shows that it is conceptually possible. But so what? The H₂O/XYZ story is a vehicle for introducing us to the conceptual possibility. Once we understand the conceptual possibility, it is easy enough to find nomologically possible earth bound examples. Fodor’s arguments against these examples are his arguments. Although it appears to be presented as such, I don’t see how this first point can be taken as an argument. Since, however, Fodor goes on to deal with nomologically apparently possible cases, nothing else he says depends upon this first point.

In arguing that those cases that do occur are accidents, Fodor asks us to consider what must and must not be the case in the examples we use. His examples of what he claims cannot be laws seem to me to discount the Twin examples by fiat:

It can’t be a law that the kinds among whose instances I regularly fail to distinguish when I apply C are very generally indistinguishable in respect of their causal properties. For, if they were, they would constitute a functional higher level kind, and it would be reasonable to say that C is the concept of that functional kind....It can’t even be a law that K₁ and K₂ are very generally indistinguishable in their effects on me. If they were, then, ceteris paribus, there would be a subject-relative functional kind of which K₁ tokens and K₂ tokens are both instantiations. (1994, 31)

But this is exactly the case with the H₂O/XYZ example that Fodor admits is conceptually possible.
If the contents of my state, when I apply C, depend upon external factors of which I am unaware and which have not had an impact upon me subconsciously, then the externalist point is secure. The fact that there may be no functionally relevant distinction between K1 and K2, e.g., jade and jadeite, which are very generally indistinguishable in terms of their causal properties, does not lead to the conclusion that C is the concept of a functional kind, such as 'dark green precious stone'. Furthermore, that I mistakenly take both K1 and K2 to be instances of C, certainly does not lead to the conclusion that C is a subject-relative functional kind. It's true that there may be some subject-relative functional kind, but that's not what C is and that's not what will be used in the ascription of content.

This was exactly the conceptual point of the Putnam examples. In fact it is curious why Fodor would even bother to discount the

---

*I have interpreted Fodor's "very generally indistinguishable" as being more inclusive than "impossible to distinguish" and, in doing so, I have assumed that the jade/jadeite case would be an example of the former. Perhaps Fodor's point is better understood in the stronger sense, as seems to be indicated by his statement of what he thinks his opponents would need to show: "...you would have to show that: It is nomologically possible that there are creatures for which it is nomologically impossible to distinguish between as and bs, but of which an externalist theory of content is required to say that they have the concept A but don't have the concept B." If, however, we do interpret Fodor's "very generally indistinguishable" in this way, then any arguments he attempts cannot get off the ground since we are no longer talking about kinds whose instances I fail to distinguish. If it is impossible to distinguish, then there is no sense in which there are two kinds of things.
Putnam example because of its nomological impossibility when it seems here that he is not accepting the force of the example anyhow. In the Putnam example, water(earth) and water(twin earth) are functionally equivalent but the conclusion is that they are still distinct concepts because one refers to H\(_2\)O while the other refers to XYZ. It seems that the point Fodor is making here would result in simply denying that they are distinct concepts since they are functionally equivalent. If this is Fodor’s position, this would make the rest of Fodor’s book superfluous, since on any example given, one could always simply claim that the two apparently distinct tokens being applied as one concept are actually tokens of the same higher order functional concept.

Thus, although I don’t distinguish between arthritis and an aching pain, they are both part of the higher order functional concept of rheumatoid ailments. A higher order functional concept is simply a concept that abstracts from the physical or other non-functional properties of the objects and is based, rather on the functional properties. Although the two objects in front of me may be different colours and shapes, they are both corkscrews and are thus both part of the higher order functional concept of a corkscrew. And although a pain in my thigh is different from arthritis in my thigh, I don’t distinguish them because they both belong to the higher order functional concept of ‘rheumatoid ailment’ or anything that causes a rheumatoid
pain. So when I think, "I have arthritis in my thigh", the concept 'arthritis' means 'rheumatoid ailment'. This is the higher order functional concept. And if I don’t distinguish between lampposts and cars, they are part of the higher order functional concept of things made with metal.

Of course in this last example the things I fail to distinguish are very generally distinguishable, but if I fail to distinguish them then they are very generally indistinguishable in their effects on me, making C, according to Fodor, a subject-relative functional kind. This point of Fodor's would be fine if he actually held and defended an individualist view of semantics. But he does not. The view he holds is informational semantics, about which he says that it is closely associated with the idea that semantical properties are externalistic and broad and that, according to broad theories, the basic semantic properties of thoughts are truth and denotation. Unless one is going to argue against the possibility of these types of twin cases on the basis of an individualistically based semantics, I don’t see how these cases can be ruled out.

According to informational semantics, according to Fodor, jade thoughts, for example, are about jade because they are the kinds of thoughts that jade can be relied upon to cause. Now, if

9 Actually Fodor uses the example of dogs in his explanation, but I wanted to use an example here that more easily illustrates an externalist point. The point could be made with dogs, in a case
jadeite can also be relied upon to cause jade thoughts then, according to informational semantics, the meaning of 'jade' must include jadeite. Arthritis thoughts are about arthritis because they are the kinds of thoughts that arthritis can be relied upon to cause, but if any rheumatoid ailment can be relied upon to cause arthritis thoughts then the meaning of 'arthritis' must include rheumatoid ailments. One final example: Water thoughts are about water because they are the kinds of thoughts that water can be relied upon to cause, but if XYZ also can be relied upon to cause water thoughts, then the meaning of 'water' must also include XYZ.

The problem with all of this is the same problem that Putnam pointed out in his example, i.e., the meaning of 'water' does not include XYZ, the meaning of 'arthritis' does not include all rheumatoid ailments, and the meaning of 'jade' does not include jadeite. If informational semantics is going to be sensitive to truth and denotation, then this way of describing it won't do. But it is only if it is described this way, i.e., in terms of its causal role in affecting speakers (or thinkers) that it can be used to argue that C constitutes a higher level functional kind in cases where one regularly fails to distinguish between kinds. If it is the causal role of a kind in terms of its effect on thinkers that determines the meaning of the concept for that

where someone's dog thoughts are reliably caused by wolves, for example, but I think the example of jade is more straightforward.
kind then we won't be able to make the distinctions that
externalist examples, such as those above, claim that we do
make. Causal role doesn't exhaust semantic role, at least
according to our examples.

So, we want to claim that it is a law that the kinds I regularly
fail to distinguish when I apply C are generally
indistinguishable in respect of their causal properties or, at
least, in respect of their effects on me, and although they may
constitute a higher level functional kind, C is not the concept
of that functional kind. It seems to me, in fact, that it is not
possible to deny that it is a law. If two things are
indistinguishable in terms of their effects on me, including the
effect of causing me to believe they are indistinguishable,
then, by stipulation, I will fail to distinguish them. For if I
distinguish two things, then my distinguishing them is at least
one respect in which they have differentially affected me. I can
only distinguish between two things that affect me in such a way
as to allow me to distinguish between them. How can this not be
a law? If it's true that twin cases require that this is not a
law then, indeed, there are no twin cases. But in this case we
would deny the existence of twin cases not because of anything
relating to higher level functional kinds, but simply because
accepting the existence of twin cases would result in the denial
of a truism. Since I see no reason to deny that it is a law,
however, I see no rationale for ruling out twin cases.
Following his stipulations of what can't be laws, Fodor claims that there still are some cases of twins, but that since these cases don't follow the laws, they are to be regarded as accidents, and as such, they don't impugn his theory. Although we have already rejected Fodor's two stipulations relating to causal properties, I think it is worthwhile to become clearer on Fodor's accident argument.

Fodor considers the jade/jadeite case to be ruled out because for those who fail to distinguish jade from jadeite such failure is accidental. Let's imagine twins named 'Johnny A' and 'Johnny B'. Johnny A has encountered only jade and Johnny B has encountered only jadeite yet when they think, "What a fine specimen of jade" they are in exactly the same physical states. Since, however, Johnny B has only encountered jadeite, when he uses the term jade he refers to jadeite, while Johnny A refers to jade. But, according to Fodor, the fact that Johnny B has only encountered jadeite and that Johnny A has only encountered jade is accidental and we don't want our psychology to capture accidental generalizations. Since there is no law that says you can't distinguish jade from jadeite, it's an accident that they don't.

The problem is that jade and jadeite are very generally indistinguishable in respect of their causal properties and certainly they are indistinguishable in terms of their effects
on the Johnny twins. If we do think that it's a law that when we fail to distinguish things it is because those two things are generally indistinguishable, then the case of jade and jadeite is an example of this law, not an example of an accident. Furthermore, while it may be accidental that Johnny A has only encountered jade and Johnny B has only encountered jadeite, this fact is not relevant to the outcome of the story. Imagine that both twins have encountered samples of both jade and jadeite. Johnny A is viewing a sample of jadeite and thinks "what a lovely specimen of jade", while Johnny B is viewing a sample of jade and thinks "What a lovely specimen of jade". The twins are in the same physical state, but one has made a true statement while the other has made a false statement. This outcome does not depend upon anything accidental.

For Fodor, if two distinct kinds of things are taken to be one kind of thing it cannot be because of anything to do with the causal properties of those things since it is on the basis of the causal properties that kinds are determined. Informational semantics takes the contents of one's concepts to be constituted by one's dispositions to apply them. One's dispositions are determined by the causal properties of the things in question. Thus, if it is not because of the causal properties and there is still a distinction drawn between the two things--there's no law that says you can't distinguish them--the distinction is
accidental. It seems to me that Fodor has not actually argued for this conclusion.

Fodor assumed the truth of informational semantics at the outset and using informational semantics stipulated that it can’t be a law that when I fail to distinguish between two things, that those two things are generally indistinguishable in their effects on me. Thus, when I fail to distinguish two things, those two things are actually distinguishable, and my failure is an accident. He takes the accidental part to follow. If it follows from anything it follows from his stipulation. Even then, even if we accept his stipulation, it seems to me that it doesn’t necessarily follow. If the two things weren’t actually distinguishable somehow, there would be no failure to account for. In order for it to be the case that we can say that I failed to distinguish between two things, there must be some sense in which those two things are distinguishable. The fact that those two things are somehow distinguishable does not entail that my failure is an accident. We can’t explain the failure away by calling it an accident. Our theories do need to account for these failures.

The problem, as I have presented it, is primarily with the stipulation that when I fail to distinguish things it is not because they are indistinguishable in their effects on me. In fact I argued that this is a truism. And even if we grant this
and agree that there it is another sense in which the things are distinguishable, it doesn’t follow that my failure is an accident. I don’t know how big a problem this is for informational semantics in general, but I think it is a problem for Fodor’s attempt to discount twin cases.

At this point Fodor takes himself to have eliminated the difficult twin cases, but cases involving deferential concepts, where we rely on experts to tell us the meaning, remain to be reckoned with: "...when I think elm I am in a computational state that is, to all intents and purposes, functionally indistinguishable from the one I’m in when I think beech....Broad content individuation insists, however, on distinguishing between these states since they have, by assumption, different truth conditions" (33-34). In cases like these, where we don’t have the expertise to tell the difference between kinds of things, we defer to experts to help us out. For these cases Fodor’s argument is straight forward. He claims that we in fact can tell the difference between a beech and an elm and the way in which we tell the difference is by consulting the experts. Most of the time we don’t care whether we can tell the difference but when we do care we consult the experts.

I don’t see how this is an answer to our problem. The problem is that if we want to maintain psychological realism, individual supervenience must be true and it is hard to see how
supervenience can be true if externalism is true. The beech/elm case illustrates one instance where it appears that the semantic properties of my belief state cannot be supervenient upon my physical state because I lack information about the semantic properties. The fact that I can, if I am so inclined, seek out this information and become able to make this distinction does nothing to change the fact that the contents of my belief state when I am unaware of the distinguishing properties does not supervene upon my physical state. Fodor suggests that this is irrelevant, however, since if I don’t care about the distinction, the fact that I don’t make it will not have an effect on my behaviour. Since we are concerned about behaviour in psychology we need not be concerned about these cases.

I think Fodor may be right insofar as he limits his point to cases we don’t care about. Fodor is agreeing that there are cases in which individual supervenience is false, but these cases are irrelevant to psychology. It seems to me, however, that one may care about whether he or she has made the right distinction and still be wrong. Furthermore, when one does care and consults an expert, doing so does not thereby render the person’s concept supervenient upon his or her physical state.

In the example given by Burge that we reviewed in the last chapter Daniel fails to distinguish the pain in his thigh from arthritis. His reason for seeking out a doctor was not that he
wasn't sure what he had. He thought he knew what his problem was and he sought out an expert for help with his problem. Perhaps he might have even begun treating the pain in his thigh with his arthritis medication. Thus, although Daniel couldn't tell arthritis from a general rheumatoid pain, it did matter to him and it affected his behaviour. In this case Daniel's belief that he has arthritis in his thigh does not supervene upon his brain state since he can't distinguish between arthritis and a general rheumatoid ailment, and if we suppose he took arthritis medication, his behaviour is affected by his belief. Had he been able to make the distinction he wouldn't have taken the medication.

Once Daniel has spoken with the doctor and he realizes that he doesn't have arthritis in his thigh he stops taking the arthritis medication. However, he still may not fully understand what 'arthritis' is. In fact it is unlikely that he will ever know as much as the doctor about arthritis and while he may learn that arthritis cannot occur in one's thigh, he may still believe that it could occur in one's stomach. Consultation with an expert does not change the fact that the content of Daniel's beliefs about arthritis is partly determined by facts that he knows nothing about and that are in no way physically represented within him. The expert may help to make Daniel behave more rationally in light of his beliefs by revealing further details about his beliefs, but the supervenience of
semantics upon external factors does not shift to internal factors simply because the subject gains more information. It seems to me that the deferential cases remain good examples of the externalist point against individualistic psychology. And it seems to me that the twin cases that Fodor attempts to rule out as accidents cannot be dismissed in the manner he suggests. Thus, although Fodor at this point plows on to consider the problems posed to his view by Frege type cases, I believe that he hasn't sufficiently shown that we don't have to worry about twin cases and that is enough to grant that externalism is still a threat to psychological realism. Unless, of course, Burge can tell us why it isn't.

Burge characterizes individualism as follows: "According to individualism about the mind, the mental natures of all of a person's or animal's mental states (and events) are such that there is no necessary or deep individuative relation between the individual's being in states of those kinds and the nature of the individual's physical or social environments." He then claims that this view is mistaken and that "...a person's intentional states and events could (counterfactually) vary, even as the individual's physical, functional (and perhaps phenomenological) history, specified non-intentionally and individualistically, is held constant" (1986, 4).
While Fodor and I have assumed that unless externalism or anti-individualism about content is false or somehow not relevant, we cannot be realists about psychology, Burge assumes psychological realism, points out that explanations in psychology are non-individualistic and argues that there is no reason to be revisionistic about psychology. Since it is not my purpose here to challenge the externalist point, but to determine whether it can be compatible with realism, let's grant Burge’s claim that one’s intentional state could vary while internal conditions are held constant. We have already reviewed Burge’s arthritis example which he takes to be an illustration of this point. What is at issue then is whether Burge’s claim that psychology as currently practiced as non-individualistic warrants the conclusion that realism is compatible with non-individualism. If we accept that the language used in psychology, while a refinement upon common sense discourse, nevertheless maintains a link to that discourse and is not purely individualistic, we will accept Burge’s claim that individualism must be revisionistic about the language of psychology. And then if we are to revise the language, we must have good reasons, showing why it should be individualistic.

The problem with this line of reasoning, it seems to me, is that even if we can’t revise psychology in the appropriate manner without our revisions amounting to what might appear to be an elimination of psychology altogether (in favour of
neurophysiology, for example), this does not demonstrate that realism is thereby warranted. That is, if it is true that psychology as currently practiced is both useful and non-individualistic and that revising it in the appropriate individualistic manner would weaken its usefulness (or render it obsolete, replaced by neurophysiology), this does not in itself entail realism about intentional states. All it entails is instrumentalism, the view that although beliefs and desires don't really exist, it is useful or instrumental for us to continue to act as though they do. But it is clear that Burge wants more than this. In discussing the use psychology makes of interpreted that-clauses and other intentional constructions, Burge says: "I have seen no sound reason to believe that this use is merely heuristic, instrumentalistic, or second class in any other sense" (1986, 8).

Instead of explicitly arguing for the compatibility of realism with anti-individualism, Burge claims that the onus is upon the individualists to show why we should accept their view. However, even if we accept that externalism or anti-individualism is right when we are considering intentional state contents and it reflects current practice in psychology, the force of the individualistic point concerning supervenience is no less threatening to realism. One need not argue for revising psychology in the face of the individualist's point. One can accept the point and deny realism. Or one can simply acknowledge
the force of the point and not have an opinion on what to do about current practice in psychology. It is a separate issue. Burge instead denies the force of the point and therefore does not deny realism. The individualist's point is put this way by Burge: "So if propositional attitudes are to be treated as among the determinants of behaviour, they must be taken to supervene on brain states. The alternative is to take propositional attitudes as behaviourally irrelevant." And his response: "This argument can, I think, be turned on its head. Since propositional attitudes are among the determinants of our 'behavior' (where this expression is as open-ended as ever), and since propositional attitudes do not supervene on our brain states, not all determinants of our 'behavior' supervene on our brain states" (1986, 13).

The problem for Burge, however, is that if we individuate propositional attitudes, not on the basis of their causal properties, but on the basis of their relations to environmental (including social) conditions, it is possible that we may not be individuating the 'thing' that is causally responsible for behaviour, which is what psychology seems to require. How can something be a determinant of our behaviour unless it is causally efficacious? According to Burge, we are picking out the thing that is causally responsible for behaviour, it's just that how we pick it out is on the basis of environmental conditions.
Burge offers an analogy based upon breathing and the state of the lungs. He points out that although we would grant that nothing causally affects breathing except as it causally affects local states of my lungs, the sub-events of respiration are not individuated in such a way that they are supervenient on the chemically described events that compose them: "If the same chemical processes (same from the surfaces of the lungs inside, and back to the surfaces) were embedded in a different sort of body and had an entirely different function (say, digestive, immunological, or regulatory), we would not be dealing with the same biological states and events" (1986, 16). Thus, although he grants that nothing causally affects my mental state except as it causally affects states of my brain, my mental states (propositional attitudes) are not individuated in such a way that they are supervenient upon the neural firings in my brain. My propositional attitudes, like the sub-events of respiration, are individuated partly on the basis of environmental or functional factors and, so the reasoning should go, this makes them no less prone to causal effects or no less causally efficacious themselves. In Burge's thought experiment then, it is not the case that if we moved Daniel-a from Earth to Twin Earth his mental state would be causally affected. It simply would get individuated differently and the differences are accounted for in terms of the external environment.
Psychology assumes that propositional attitudes are legitimate means of explaining behavior. And it assumes that the manner in which that mental state responsible for behavior is picked out is essential. Thus although ascribing to me the belief that today is Wednesday will serve as an explanation of my behavior of leaving my office early, ascribing to me the belief that today is Jane Smith's 45th birthday, which happens to fall on a Wednesday, will not serve as an explanation since I don't know when Jane Smith's birthday is. In the case of propositional attitudes, unlike the case of the sub-events of respiration, the way in which we individuate the state we are talking about may have an impact upon the causal powers we can attribute to that state.

The way we individuate, or the description we use to pick out, a particular sub-event of respiration is irrelevant to the causal powers of that event just as whether we pick out a particular star by 'the morning star' or 'the evening star' is irrelevant to the star's causal powers. In such cases there are objects or events that exist independently of our descriptions and when we talk about causal powers, we are interested in powers of the physical objects or events, not of the objects or events under a certain description.

Compare propositional attitudes according to Burge. The content of the propositional attitude is not individuated in terms of
its physical composition. But if we are realists about propositional attitudes, there still must be some underlying 'thing' or state that the propositional attitude descriptor picks out that has a particular physical composition. It is this state, this particular physical composition, that determines or is responsible for causal role. This state can be picked out by different descriptions, depending on the context, again just like the morning star and the evening star and sub-events of respiration. But as we just saw, unlike the case of respiration, it can't be just any co-referential description. While a particular star can be picked out by different but co-referential descriptions, it seems that when it comes to beliefs, changing the description changes the belief contents which, in at least some cases, changes the causal properties. Intentional contexts do not allow for substitution of co-referential terms.

But perhaps Burge did not intend that any manner of individuation would be acceptable. Perhaps altering individuation by substitution of co-referential terms within the intentional context would be ruled out. So let's consider Burge's own types of examples. Recall that Burge wants to maintain that propositional attitudes are among the determinants of behaviour and that nothing causally affects my mental state except as it causally affects states of my brain. So although changing contexts may require us to change how we individuate a
particular mental state (whether we attribute 'arthritis' thoughts or 'tharthritis' thoughts) nothing has causally affected my brain and nothing has causally affected my mental state. The labelling of the mental state and the content of the propositional attitude has changed, but this is not due to anything causal, nor has it affected the causal properties of the state. The question at issue is whether it is true that beliefs that we distinguish according to content may nevertheless be indistinguishable according to their causal properties. Are the causal properties of arthritis and tharthritis the same?

The question is not straightforward. If we assume that externalism is true, as I have been, then we are in agreement with Burge that Daniel-a and Daniel-cf can be in the same brain state and have two different beliefs and this means, of course, that the causal properties would be the same. On this reasoning, however, the causal properties are the same by stipulation. If externalism is true, the causal properties are the same. But since we are trying to determine the compatibility of externalism with realism about psychology, we must consider how psychology would determine the causal properties.

On Burge's example, Daniel-a and Daniel-cf go to see their doctors and report their fears that their arthritis has spread to their thighs. In order for Burge's rejection of the
individualist point to be secure we should find upon analysis that psychology would attribute the same causes to Daniel-a’s and Daniel-cf’s behaviour. Recall that Burge takes them to be believing different things. Daniel-a believes that he has arthritis in his thigh while Daniel-cf believes that he has tharthritis in his thigh. Furthermore Daniel-a is incorrect in his belief while Daniel-cf is correct. It seems to me that the explanation for Daniel-a’s behaviour of going to see the doctor would be that he believed that he had arthritis in his thigh while the explanation for Daniel-cf’s behaviour of going to see the doctor would be that he believed that he had tharthritis in his thigh. If this is the case and psychology does offer different explanations for the behaviour, then psychology is, it seems to me, committed to there being different causes of the behaviour. Tharthritis, for example, would not work as an explanation of Daniel-a’s behaviour and the belief that he had arthritis in his thigh would not work as an explanation for Daniel-cf’s behaviour. Thus substitution of co-referential (as assumed by the example) terms does not work for Burge’s example either. It doesn’t work because when psychology offers a particular belief as the cause of a behaviour it doesn’t, it seems to me, take itself to be offering a description of some underlying state distinct from the content of the belief. Psychology is not the science of describing brain states but is supposed to be the science of behaviour. If the descriptions given by psychology are not causally relevant psychology cannot
be the science of behaviour. Thus, since the psychological explanation of Daniel-a's behaviour is different from that of Daniel-cf's behaviour, psychology must take them to be causally different.

Although, as per Burge's example, our individuation of the respiratory system depends partly upon environmental conditions and we can list a variety of individuating descriptions to pick out the same physical system embedded in its environment, in the case of beliefs, realism in psychology requires more than this. In the case of psychology we don't simply pick out a belief with a description. When it comes to psychological states like beliefs and desires, we are positing the state at the same time that we are describing it. We are advancing an hypothesis that the belief exists at the same time that we are saying what the belief contents are. This, at least, is what we do according to the widespread view that folk psychology is a theory and that beliefs are posits in that theory. The only evidence we have for the existence of the belief is the behaviour that it is hypothesized to explain.

With the respiratory system and our individuation of the lungs, we can look and point at the lungs and ask what this physical thing here, that I can see, is for. We can then decide to refer to the lungs in terms of causal role or in terms of the lung's location in the body. The scientific approach to individuating
the lungs is concerned with causal role which is specified in terms of relational factors. What is different about beliefs is that there is nothing that we can point at and say there is the belief and here's how we are going to individuate it. Since we can't directly observe beliefs, the individuating or describing and the pointing come to one and the same thing in the case of beliefs. In other words, if we accept that folk psychology is a theory and that beliefs are posits in that theory, the individuation and the determination of beliefs come to the same thing. Belief contents cannot be relative to context because belief contents do not simply individuate beliefs, they determine them. If psychology takes beliefs to cause behaviour and beliefs are not individuated by descriptions but are determined by those descriptions, relativism with regard to the descriptions is not allowed. If you change the description, you change the belief, and if you change the belief, you change the cause. Again, however, if we take psychology to be the science of describing brain states, then the analogy with the respiratory system may go through more smoothly. But this approach, it seems to me, begs the question.

Assume that we are describing a brain state when we say of Peter that he believes that Mickey is a mouse, and that that brain state, whatever state it is, is causally responsible for Peter's buying cheese for Mickey rather than, say, dog food. Twin Peter has a different belief. He believes that Mickey is a twmouse and
is caused by this belief to go and buy some cheese. If we are to accept the analogy with the respiratory system, the brain state that is identified with Peter's belief is the same as the brain state identified with twin Peter's belief, so the causes of the cheese-buying behaviour are the same. We are simply describing the causes differently. The problem is that we don't know, in first place, when we ascribe the belief that 'Mickey is a mouse' to Peter, what, if any, brain state we are describing. We have no other description of it. We have no other way of individuating it. How can we know, then, that a different belief, twin Peter's belief, describes the same thing? The response that they have the same causal powers is blocked because, since we haven't individuated the beliefs on the basis of their causal powers\textsuperscript{10}, the only way we can know that they have the same causal powers is to assume that they describe the same brain state\textsuperscript{11}. But that is what we are trying to prove.

Realism in psychology requires, at least on the model where folk psychology is a theory, that when we list a variety of beliefs

\textsuperscript{10} Even though we have ascribed the belief to Peter that 'Mickey is a mouse' and to Twin Peter that 'Mickey is a twmouse' because we wanted to explain their cheese buying behaviour, remember that, according to Burge, the beliefs aren't individuated on that basis. They are individuated, in part, by their social and environmental context. That is why they count as different beliefs.

\textsuperscript{11} The fact that Peter and Twin Peter both buy cheese does not sustain the conclusion that they are in states with the same causal powers. Cheese-buying behaviour can be caused by many different states. The only way the view that they have the same causal powers can be sustained is if we already know that they're in the same brain state.
we are listing a variety of different explanations and when we list a variety of different explanations, we are thereby identifying different causes of behaviour. When Daniel-a went to the doctor we might explain this behaviour by saying that he believed he had arthritis or that he believed he had tharthritis or perhaps that he believed he had the flu, but in doing so we are offering competing explanations and competing causes. Psychological realism assumes, it seems to me, that when we offer different explanations we are individuating different causes, not that there may only be one state which can be denoted by a variety of beliefs (depending on context). For the entities posited by psychology to be real, only one of these explanations can be right. Realism about psychological states requires that there must be a fact of the matter about what Daniel believes.

Burge would not disagree that psychology would interpret Daniel-a and Daniel-cf as holding different beliefs. In fact this is his own position. Since he believes that psychology is non-individualistic, the context would be relevant to the belief attribution. He would disagree, rather, with the claim that the explanations being different entails that the underlying causes are different.

When psychology offers a belief as an explanation of behaviour, does it take itself to be offering the cause of the behaviour or
to be offering a context-relative description of a state which is causally responsible for the behavior? Is it compatible with realism about psychology to agree with Burge? He says that: "The same physical transactions in a given person may in principle mediate, or underlie, transactions involving different intentional states—if the environmental features that enter into the individuation of the intentional states and that are critical in the explanatory generalizations that invoke those states vary in appropriate ways" (1986, 17). Burge's view is that the claim, that none of an individual's beliefs could have been different unless his brain states were different, is not a claim about what is required for realism in psychology and, in fact, is nothing more than metaphysical conjecture.

While Burge does argue that individualism is false he does not offer any arguments that the denial of supervenience is compatible with realism in psychology. He simply claims that since psychology as practiced is non-individualistic, the onus is on others to demonstrate why psychology should be revised. If psychology requires the assumption that the explanations it offers for behaviour are the genuine causes of behaviour and not context-relative descriptions that may vary while the causal properties of the underlying states to which they bear a relation stay the same, then psychology should be revised. If psychology as practiced does not require the above but realism about psychology requires it, then perhaps psychology should not
be revised. Although its explanations would not reflect the truth about the non-existence of the entities it posits, it may still be the most instrumentally effective way to do psychology. It may be the only way to do psychology.

Whether realism survives does not depend upon what psychology looks like now or how it is practiced. It depends upon what assumptions are required for realism. Psychological realism, on the face of it, requires, it seems to me, that the explanations in psychology are to count as denoting the causes of behaviour, and causes cannot be context-relative. Psychological realism requires that either I went to the doctor because I believed that I had arthritis or I went to the doctor because I believed I had thartritis. If my belief is to count as the explanation of my behaviour, psychology cannot change its explanation with context. If the psychological explanation does change with context, then the psychological explanation must be leaving something out. It does not reveal that the cause of behaviour hasn’t changed and it’s not clear how it refers to or individuates the cause of behaviour.

Typically realism and relativism are understood to be alternative positions. It appears to be Burge’s contention that although explanation in psychology is relative, we can nevertheless maintain that the objects posited by those explanations are real. I agree with Burge’s contention that our
explanations are socially and context relative and I believe that academic psychology is a successful enterprise and that we cannot get along without the assumptions of folk psychology. Unlike Burge, however, I still feel the need for an argument to the effect that realism is true despite the relativism of explanation. Putting the burden upon the individualists to propose a revised psychology misses the point. The burden is upon Burge to show just how it can be the case that our context-relative explanations do refer to the causes of behaviour.

Like Burge, Stephen Stich (1996, 168-91) thinks that the demand for the truth of supervenience for the defense of psychological realism is misguided. Stich’s starting point, however, is quite different from both Fodor and Burge. Burge doesn’t deny the physical realization of intentional states. His thesis is that we simply pick out this physical/intentional state with a variety of different descriptions. Fodor thinks that denying supervenience results in denying realism about the intentional because the belief contents must themselves be physically represented and have causal properties. It seems that Fodor, along with many other philosophers, believes that the intentional must be naturalized or intentional irrealism will result. Stich agrees that intentional irrealism is a very radical doctrine. He claims that, if it looked like anything else we held led to intentional irrealism as a conclusion, we should, in fact, be very concerned about intentional irrealism.
But, Stich claims, there is no reason to think that the failure to naturalize the intentional will result in intentional irrealsim.

Stich rejects all forms of supervenience. Thus, not only does he agree with Burge that individual supervenience is false but he also rejects global supervenience. The claim of global supervenience is that, given another world exactly the same as this one in all physical respects, it will be the same in all intentional or psychological respects. This he thinks is false because he thinks that historical properties matter to the determination of intentional state contents. Thus, although my doppleganger on Twin Earth can be in exactly the same bodily physical state as I’m in and be situated in exactly the same world in terms of physical descriptions, we still are not having thoughts with the same contents when we each think "Julius Caesar was Roman". This is because we can imagine that Twin Earth came into existence only moments ago and, thus, while my Julius Caesar thoughts are causally linked to Julius Caesar my doppleganger’s are not.

After taking himself to have established that global supervenience is false for intentional properties, Stich then claims that it doesn’t matter anyhow since realism doesn’t depend on the truth of supervenience. He points out that there are many properties whose realism we wouldn’t deny but that
don't supervene on physical properties. He cites being a Picasso and a genuine $100 bill as examples (1996, p.187). Even though something may have the same physical properties as a Picasso on our newly created Twin Earth it's not really a Picasso because it wasn't painted by Picasso and the $100 bill cannot be genuine since part of the notion of being genuine is that it come from the right place. We don't conclude from this that there are really no Picassos or $100 bills. In the same way, we shouldn't conclude from the denial of supervenience that there are no beliefs.

It seems to me that Stich's analogies are not, in fact, reassuring about the lack of threat posed to intentional realism by the failure of supervenience. We have never doubted realism about Picassos and genuine $100 bills. We need to ask why we would ever call into question realism about intentional states but not realism about Picassos. Stich's analogies don't demonstrate that supervenience is unnecessary for realism unless we already grant that realism about the intentional depends, at least in part, on properties that are not causal properties.

It certainly seems right to say that part of what makes a Picasso a Picasso is its causal properties. The power of the physical arrangement of lines and colour on a canvas to evoke certain emotions or thoughts in people is apparently a significant factor in our determining whether something is a
Picasso. But having the causal properties of a Picasso is not sufficient to being a Picasso. How the thing with those causal properties came to acquire those causal properties is actually the only determining factor of whether that thing is a Picasso. Something can only be a Picasso painting if it was painted by Picasso. Being painted by Picasso is itself a sufficient condition of being a Picasso, even if the painting is one that does not have the same or similar causal properties to other Picasso paintings. So not only are the typically associated causal properties not sufficient, they aren’t even necessary. A bad Picasso is still a Picasso. Thus, the failure of a Picasso to supervene on physical properties is not surprising. And the same can be said for the genuine $100 bill. If, for some reason, inflation in our economy ran wild tomorrow, the same $100 bill would have distinctly different causal powers tomorrow than it has today. While today I can purchase a week’s worth of groceries for my family with it, tomorrow it will only allow me to feed myself for a week. Regardless, there is no question about the genuineness of the $100 bill. Again, the failure of the $100 bill to supervene on physical properties is not surprising.

The failure of mental states to supervene on physical states would not be surprising if, like with Picasso paintings and $100 bills, the causal properties of mental states were irrelevant to their existence. If the causal properties of the belief that
'Santa is coming to visit on Christmas eve' were different, would it still exist as the same belief? In the context of intentional states, it seems like an odd question to ask whether we can change the causal properties without changing the belief.

The externalist examples demonstrated that we can change the belief without changing the brain state, which may mean that we can change the belief without changing the causal powers. Burge didn't take this to imply that beliefs were causally irrelevant. He took it to mean that which beliefs individuate the right causes will vary with context. The question now before us is whether the belief can remain the same if we change the causal properties associated with the belief. Presumably this is not something that Burge would accept. Accepting this latter point would be to accept that the causal properties of beliefs are irrelevant. If we accept this, it seems to me that we can have no science of beliefs and, as we have seen, Burge thinks that psychology is both anti-individualistic and scientific. It seems to me that anyone who wants to maintain psychological realism cannot, in fact, accept that the causal properties of beliefs are irrelevant. This being the case, perhaps we should consider whether Stich really does accept this claim.

Stich does, in fact, acknowledge the objection and grants, "for arguments sake", that intentional properties are not causally
efficacious: "Would this be a major catastrophe? So far as we can see, it would not be a catastrophe at all. For, given any intentional property, it is easy to find a "narrow" surrogate of that property which does supervene on the current, internal, physical state of the organisms" (p.183). Ultimately then, Stich ends up back where we began in the last chapter, advocating narrow content. As I argued then, it seems to me that once we move to narrow contents, we are no longer talking about psychology as currently practiced and that since no reasonable notion of content can ignore historical and environmental factors, the narrow content proposal cannot be accepted as a proposal about intentional psychology.

Most philosophers worry about the causal properties of intentional states because without them we could have no laws and, hence, no science of psychology. Stich, however, argues separately that the failure of supervenience does not imply that intentional properties cannot play a role in laws. Using the analogy of the HIV virus, Stich points out that the evolutionary ancestry of the virus is essential to its being the virus. So, the HIV virus is another example of something about which we have no doubt of its existence, yet it doesn't supervene on its current physical properties. He argues that it is likely that there are laws concerning the behaviour of the HIV virus even though it doesn't supervene on its current physical state:
Though the details are still to be worked out, it is plausible to assume there is a law-like connection between infection by the HIV virus and the death of certain cells that play an important role in the immune system. ...But if the current worry were correct, then there could be no such law, because being infected by the HIV virus is not a property that supervene's on an organism's current, internal, physical state. ...Analogously, the fact that intentional properties do not supervene on the current, internal, physical states of organisms does not entail that intentional properties cannot play a role in laws. (1996, 184)

The problem with this analogy is the same as with the Picasso and the $100 bill examples we examined earlier. In the case of the HIV virus, its existence does not depend on its causal powers. The fact that we now have medication which allows some people to live ten to fifteen years longer than was originally expected does nothing to our claim that those people have the virus. The HIV virus currently causes illness and death. In the future, hopefully sometime in the near future, the virus will no longer have these effects on people. Nevertheless, there will still be people who have the virus. The existence of the virus depends on its history, not its causal powers. Again, this is not the case with beliefs. If we take away the causal property of a belief, its not clear what we are left with. The standard view of beliefs is that they are posited to explain behaviour. Beliefs are understood to be states with certain causal properties. If those causal properties change and beliefs no longer explain behaviour, either we must change our standard view of beliefs as causal entities posited to explain behaviour, or we must reject realism about beliefs.
But the point of Stich's example in this case is not simply that the HIV virus is another example where we don't deny realism because of the failure of supervenience. It was intended as an example of a case where even though supervenience fails, we can have laws. The problem is that if there were a law that connected the HIV virus with the death of certain cells, the causal properties associated with that law would supervene on the current physical state of the organism. Assume, for example, that John has the virus and John's doppleganger on Twin Earth has a virus that is physically exactly the same as John's. We are in agreement that we can't call it the HIV virus, because we are assuming that Twin Earth just came into existence a few moments ago and, thus, the virus that Twin John has does not have the history to be called the HIV virus. Nevertheless, if we knew that Twin John were in exactly the same current physical state as John, we would expect his virus to behave in the same way as John's. If we found that while in John's case the infection seemed to cause the death of certain cells, yet those same cells remained alive in Twin John, we would not accept that it is a law that the virus causes the death of certain cells. 12 The non-physical, social and historical factors associated with the HIV virus do not enter into any laws we may discover.

12 This, of course, assumes that the environment and all other physical conditions on Twin Earth are exactly the same as on Earth. We mean to rule out possibilities such as the air being of a quality such that the cells don't deteriorate in the same way.
Stich offers another example. He cites Greshem’s law, which is a law of economics. The law claims that bad money drives good money out of circulation. Stich argues: "Plainly, neither the property of being money nor the properties of being good and bad money supervene on the current, internal, physical state of coins, banknotes, wampu, and the like. But this is no reason to suppose that Greshem’s law is mistaken" (p.184). This example does not, like the HIV virus, depend on the law drawing only on the current physical properties of money. Whether there are actually any laws in operation in economics, however, is at least as contentious as whether there are any laws in psychology. In fact, laws of economics are based on human behaviour with regard to money. Stich’s example thus begs the question. We can’t cite economics as an example of a case where laws exist that do not supervene on physical properties to demonstrate that psychological laws need not supervene on physical properties, since economics depends upon psychological ‘laws’ or generalizations in the first place.

Stich thinks that our worries about intentional realism are misplaced. He thinks that the failure of supervenience does not pose a problem for realism. About this I agree with Stich. The analogies he presents, however, do not support his claims. I think that the problem runs deeper than Stich seems to recognize. The failure of Stich’s analogies, the fact that he resorts to narrow content and the fact that there are so many
philosophers who do think that the failure of supervenience is a serious problem is, it seems to me, evidence to support the idea that the problem is a deep one. It also seems to be one for which the responses or intuitions vary widely.

It is, I think, a good illustration of how far apart intuitions can be on this matter that Burge argues for realism by claiming that psychology is externalistic through and through, Fodor attempts to maintain realism by minimizing the force of externalism in psychology, and Stich thinks that the whole debate is unnecessary. It is also an illustration of how strongly we want to be able to maintain realism. I believe, however, that the motivations behind our unwillingness to abandon the physical reality of beliefs and desires can be reconciled within an alternative framework. At least this is what I will argue next.
Chapter Three

Realism, Instrumentalism, and the Theory of Folk Psychology

According to the arguments of the last chapter, the attempts by Fodor, Burge and Stich to save realism didn’t work. In this chapter and the next I will argue that we can, in fact, maintain psychological realism and that Burge and Stich, but not Fodor, were at least partially right after all. The trouble with the arguments presented by Stich and Burge was not with the form of their arguments, but with their failure to situate those arguments within the right framework. The right framework, I will argue, is one that rejects the idea that folk psychology is a theory and that intentional states are posits in that theory. It is because of the assumption that folk psychology is a theory, I will argue further, that concerns are raised about realism in the first place. Concerns about realism of the psychological states like beliefs and desires should, at least on the face of it, be distinguished from concerns about causal efficacy. But, aside from Stich and Burge, philosophers do tend to assume that the denial of supervenience leads to the denial of causal relevance which, in turn, leads to the denial of realism. It is only with the assumption in mind that folk

psychology is a theory, I will argue, that we find the denial of supervenience so devastating to realism.

What I will be arguing for, then, is the adoption of an alternative conception of folk psychology. I will begin by arguing that in addition to the force of the externalist arguments against realism there are further reasons to re-examine our whole conception of folk psychology as a theory. These reasons will involve a discussion of the actual possibility of rejecting the existence of beliefs, a comparison of our approach to the problem of the existence of belief with an approach to the problem of induction, and an analysis of the notion of 'behaviour'. I will argue in the next chapter that when we adopt an alternative conception of folk psychology, wherein the reality of the states is assumed rather than posited, the causal efficacy of the states is not undermined by externalist concerns. Let's begin, however, with a review of the notion that folk psychology is a theory.

What is meant by 'folk psychology' is a set of laws or generalizations of which we all possess a command or, at least, the capacity for command. These general statements have to do with mental states and their interactions with behaviour and with one another and include things like the following: persons who are smiling tend to be happy; persons who are angry tend to be unhappy; persons who are hungry tend to make attempts to
obtain food; persons who desire that P and believe that Q will result in P, will attempt to bring it about that Q. Folk psychology, viewed as a theory, is a theory of behaviour. It posits mental states to explain behaviour. States like being angry or believing that it's raining are posited to explain the behaviour of yelling or grabbing your umbrella.

According to some philosophers, the existence of these states, if they are theoretical, depends upon the truth of the theory. If the theory is false, there is no reason to maintain that these states exist. This is the position held by the eliminative materialists. Thus the eliminativists foresee a much better framework, the framework of neurophysiology, which will posit states with much greater explanatory force than states like beliefs and desires. And since beliefs and desires are theoretical posits and nothing more, they will be eliminated and replaced with this better framework. Quine, for example, puts it this way:

If there is a case for mental events and mental states, it must be just that the positing of them, like the positing of molecules, has some indirect systematic efficacy in the development of theory. But if a certain organization of theory is achieved by thus positing distinctive mental states and events behind physical behaviour, surely as much organization could be achieved by positing merely certain correlative physiological states and events instead. (1960, 264)
Although eliminative materialism is not a currently popular theory, the view of the structure of folk psychology that eliminativists espouse, i.e., that it is a theory and that beliefs and desires are theoretical entities posited within that theory, is commonly held. This means that when we ascribe beliefs and desires to others to explain their behaviour we are applying a theory. It is this standardly held view, I maintain, that is responsible for the view that externalism threatens realism. This view is a third person view of beliefs and desires. It totally ignores first person experiences. If first person experience is discounted, or explained away as an effect of a faulty conceptual framework, then we have no source of evidence of mental states save their causal properties. Externalism, which appears to show that the causal properties are irrelevant, would thus make mental states irrelevant and, if our belief in their existence is based on their causal relevancy, there would remain no reason to continue to believe in their existence.

This is why, as we saw in the last chapter, Stich, who explicitly argues in favour of the view that folk psychology is a theory, ultimately needed to talk about narrow contents. And this is why Burge needs to give an argument to defend his position, rather than simply claiming that the onus is on those who want to revise psychology in light of the failure of supervenience. The onus, it seems to me, is on Burge to show
how, if supervenience is false, both realism and causal relevancy can be maintained. And, as I will argue, the way to do this is to reject the first assumption that folk psychology is a theory.

The view that folk psychology is a theory is not granted by everyone. The view that our ascriptions of mental states to others is based on a theory is known as the 'theory theory'. The alternative that I shall be arguing in favour of is known as the 'simulation theory'. The simulation theory claims that when we ascribe mental states to others we do so on the basis of a process that we undergo whereby we mentally place ourselves in the environmental context of the other and we 'simulate' their mental state. The simulation theory, unlike the theory theory, thus assumes the existence of mental states aside from their usefulness at explaining the behaviour of others. The simulation theory draws on our first person experiences. Before I turn to the explanation of how such a theory will enable us to maintain realism, I will consider our motivations for doing so. We will consider, that is, why a re-evaluation of our conception of folk psychology is in order.

Why don't we just accept the conclusion that appears to be a result of externalism that psychological realism is not possible? Suppose that tomorrow when you picked up your morning newspaper the headline read, "Philosophers Prove Beliefs Don't
Exist". Would you believe it? If you did, would you stop having beliefs? Would you stop attributing beliefs to others? It seems to me that no matter what evidence we are given to the contrary, we would be unable to stop understanding others in terms of their being thinking, goal directed, intentional systems. Of course this in itself does not prove that beliefs do exist. But if I'm right, and we are unable to abandon the use of the notions of folk psychology even if they were demonstrated to be false, we would be left in the very peculiar position of knowingly basing all of the interactions in our lives on fictions. Even though I know that there is no basis in my interpreting your behaviour of entering your office, for example, as intending to do some work, I could either be irrational and interpret it in this way anyhow, or reject my interpretation as illusory and not attempt to interpret your behaviour. Interpretation and prediction of behaviour, in short, would not be rationally possible.

But this still does not prove that beliefs exist. Perhaps it does indicate, however, that we are making some sort of tactical error in our manner of looking for evidence of the existence of beliefs. Rather than looking for evidence of beliefs as causal entities affecting behaviour, perhaps we should begin by granting that the way we understand one another is by attributing beliefs. Perhaps the fact that, through our attributions of belief, we do for the most part understand each
other, should be taken as some sort of justification itself for the truth of folk psychology and the existence of beliefs.

In the last chapter I criticized Burge for failing to justify realism in light of his anti-individualism. I complained that Burge had merely described what psychology does and that that doesn’t constitute a justification. Perhaps, however, a description of what we do is, in certain types of cases, enough to warrant granting the reality of the concepts involved in what we do. This, at least, appears to be the position put forward by Hume and defended by Nelson Goodman with regard to the problem of induction. Since it seems to me that the notion of belief is at least as resilient as the notion of induction, I will briefly review Goodman’s arguments here.

Goodman argued that the problem of induction that philosophers have been referring to as ‘Hume’s problem’ had actually been solved by Hume himself. In his claim that regularities in experience give rise to habits of expectation, Hume offered a descriptive approach to the problem of induction. He described what we do, but not why we do it or on what basis it is valid. Those who object to Hume’s answer to the problem of induction argue that to describe the process is not to justify it or establish that it is valid. Most philosophers saw this objection as so obvious that they interpreted Hume as simply redescribing
the problem rather than attempting to answer it. Hence, we speak of 'Hume's problem' rather than Hume's solution.

Goodman claims that there is nothing wrong with Hume's approach and that he was, in fact, solving the problem. He claims that the approach Hume took is the same approach we take in deductive logic about which no one questions our justification:

The point is that rules and particular inferences alike are justified by being brought into agreement with each other. A rule is amended if it yields an inference we are unwilling to accept; an inference is rejected if it violates a rule we are unwilling to amend. The process of justification is the delicate one of making mutual adjustments between rules and accepted inferences; and in the agreement achieved lies the only justification needed for either. (1983, 64)

This is the process we use with deductive logic and it applies equally well, according to Goodman, to induction. The problem, which Goodman illustrates through his famous 'grue' example, is not with the fact that we base our inductive predictions on past regularities, but that we cannot say which regularities are lawlike.14 "Regularities are where you find them, and you can find them anywhere" (1983, 82).

14 "Grue" is a predicate which applies to all things examined before t just in case they are green and to other things just in case they are blue. When I examine the emerald in front of me I find that it is green and this holds for all the other emeralds I examine today. Thus, by induction I conclude that all emeralds are green. The problem is that the evidence that I used to confirm the hypothesis that all emeralds are green also appears to confirm the hypothesis that all emeralds are grue. Subsequent to t, however, all emeralds cannot be grue. Although only one of the hypotheses is correct I appear to have equal evidence for both.
The problem of the justification of belief attribution seems to me to present a similar case to that of induction. If we cannot find, or if indeed it looks as if it is logically impossible (according to some externalist arguments) to find, 'entities' that match up with our notions of beliefs and desires, then, the traditional reasoning runs, we should abandon the concepts of belief and desire altogether. It seems to me, however, that we can no more abandon our practice of attributing beliefs and desires in our predictions and explanations of behaviour than we can abandon our practice of inductive reasoning. If the conclusions of our arguments don't match up with our inferences or our everyday beliefs, this doesn't necessarily imply that we should abandon our inferences or beliefs. It could imply that we should abandon the type of reasoning being used in the arguments. And just as in the case with inductive reasoning, perhaps describing the process does in itself give some sort of justification. We do use induction to make accurate predictions and we do use beliefs to predict and explain behaviour.

Burge's approach to the problem of belief ascriptions can perhaps be better understood if we treat it as analogous to Hume's treatment of induction—they are descriptions. Burge describes how psychology operates in detail and explains how it is both anti-individualistic and successful. And, as discussed in chapter one, he doesn't appear to take much notice of the fact that externalism is generally regarded as a threat to
realism and that those arguing for realism usually find themselves having to deny the force of externalism.

Regardless of whether this is the right way to interpret Burge, it seems to me that taking a description of the state of affairs when intentional state ascription takes place as a justification for the practice is appropriate. It is appropriate because, like the practice of inductive reasoning, it is not something that we can abandon. If our theory of induction tells us that it is not possible to justify induction, then perhaps we should reject the theory since we know that we do use induction and we are, in some cases, justified in doing so. The question is how do we know we are justified in those cases where we do use it. If we accept as a basic truth, rather than a theoretical postulate, that beliefs explain behaviour, the fact that supervenience is false is interesting but not threatening to realism about psychology. The truth of this fact will cause us not to abandon psychology, but to consider how we can make sense of the notion that beliefs cause us to behave in light of the fact that they do not supervene on brain states.

Many philosophers have, in fact, followed this type of route. Granting that it is practically not possible to give up beliefs and, in accordance with Goodman's view on induction, taking the fact that we do use beliefs to explain and predict behaviour as a premise to be granted rather than a conclusion to be
challenged, they have focused their efforts upon investigations of how we use beliefs and what determines their content in given contexts rather than on attempting to prove their existence. Of note is Daniel Dennett whose intentional stance view, I believe, offers a convincing explanation of the workings of psychology.\(^\text{15}\) Dennett, however, has been labelled an instrumentalist rather than a realist. He regrets this label and claims that: "My ism is whatever ism serious realists adopt with regard to centres of gravity and the like, since I think beliefs (and some other mental items drawn from folk psychology) are like that—in being abstracta rather than part of the 'furniture of the physical world' and in being attributed in statements that are true only if we exempt them from a certain familiar standard of literality" (1987, 72). Although Dennett’s intentional stance view offers an illuminating and convincing position on the context bound nature of psychological attributions, the question of realism vs instrumentalism is still one that dogs him. I believe that within the context of an alternative conceptualization of folk psychology, this would not be the case.

\(^{15}\) Dennett’s ‘intentional stance’ view is that in order to predict and explain one another’s behaviour we take the stance toward them that they have beliefs and desires. Dennett lists three principles that we follow when we approach each other as intentional systems: a system’s beliefs are those it ought to have given its perceptual capacities, its epistemic needs and its biography; a system’s desires are those it ought to have, given its biological needs and the most practicable means of satisfying them and; a system’s behaviour will consist of those acts it would be rational for an agent with those beliefs and desires to perform (1987, 49).
If it's true that we can't practically get rid of the notion of belief or stop using folk psychology then, I am suggesting, we should begin with the premise that beliefs do exist and work around problems like supervenience. It seems to me that supervenience is a problem because it is assumed that: 1) beliefs are theoretical entities posited to explain behaviour and 2) what we call 'behaviour' is itself supervenient on the physical individual. I will argue that there is good reason to reject assumption 1) and that assumption 2) is just plainly false. If one rejects only assumption 2), then while the problem of supervenience is no longer pressing, the issue of realism vs. instrumentalism remains. Daniel Dennett, for example, would agree with the first but not the second assumption. His acceptance of the first assumption is, I believe, the reason he can't escape the 'instrumentalist' label. If one rejects assumption 1) but maintains that behaviour is a physical individualistically specifiable event, then the problem of supervenience can still be overcome within the reconceptualized picture of folk psychology. But if one rejects the individualism of behaviour, weight is added to the rationale for reconceptualizing folk psychology. I will begin the discussion with the second assumption.

We have been operating with the understanding of belief as a mental state that causes behaviour. The supervenience problem arises when two different beliefs result in the same brain state.
and cause the same behaviour, since if the contents of belief have causal efficacy, then a change in belief must be reflected in a physical change in brain state. The problem with this reasoning is that the same externalist arguments that worked against individualistic construals of propositional attitude states also work against individualistic construals of behaviour. That is, what we understand as the behaviour we are attempting to explain depends, it seems to me, upon the context.

Consider, once again, Putnam’s Twin Earth example and Burge’s arthritis case. Both examples could have just as easily been used to support the view that ascriptions of behaviour are non-individualistic. Oscar1 and Oscar2 could both be in the same brain state and both be moving in the same physical manner and yet we might ascribe different behaviours to them because of the differences in their environments. They could, for example, both be walking outdoors holding umbrellas, and we could describe Oscar1 as walking in the rain while, since water is not H₂O on Twin Earth, Oscar2 would not be walking in the rain. He could be walking in the twrain, but he could not be walking in the rain. Same physical states, different behaviour. In the arthritis case, Daniel-a’s behaviour when he talks about the pain in his ankle could be described as ‘talking about arthritis’ behaviour. But even though the exact same physical motions and brain states are taking place in Daniel-cf, his behaviour cannot be described
as talking about arthritis, since there is no such thing. His behaviour is described as talking about tharthritis.

These examples, it might be suggested, intentionally taint the notions of behaviour and, in general, we can specify the behaviour in individualistic terms. In both cases, if we described the behaviour in terms of bodily motions, for example, we would be describing the same thing for each twin. Thus supervenience of behaviour on physical individuals is not violated since bodily motions certainly supervene on physical individuals. Although I think that even this claim is debatable, since what counts as a bodily motion may depend upon the context\textsuperscript{16}, I will grant it since the real problem is not whether what counts as a particular bodily motion is context relative, but whether when we talk about bodily motions, we are still talking about behaviour.

There may be cases in which the answer to a question about behaviour, what someone did or is doing, will involve nothing more than bodily motions. "She’s moving her arms up and down", may be an appropriate response to a query about what type of exercise someone is doing, for example. Generally, however, behaviour is described in terms that make reference to events or things that exist outside of the individual and, this being the

\textsuperscript{16} What counts as raising your hands, for example, is not the same when considered by a teacher in a classroom as by a police officer in the line of duty.
case, the description of behaviour is subject to the same context relative concerns that were raised in reference to belief states. In fact, it is sometimes the case that the description of behaviour and the cause of the behaviour amount to the same thing. John observes Mary get in her car and drive off and asks Sam, "What's she doing?" The reply, "She's going fishing", both describes her behaviour and gives a reason for it. She's driving off because she's going fishing or she desires to go fishing. If the behaviour can be described in the same way as the cause then obviously the same issues of supervenience apply.

This example illustrates a further problem. Since belief is often bound up in what we call 'behaviour' it is unclear how we determine which mental states belong to the description of behaviour and which ones are to count as the causes of behaviour. John enters the kitchen and sees Janet working under the sink. He asks: "What are you doing? Fixing the sink?" She replies, "No, I'm looking for my ring. I dropped down the drain." Although the bodily movements associated with 'fixing the sink' may at time t be exactly the same as the bodily movements associated with 'looking for the ring' at time t, one description of the behaviour is correct while the other is not. Whether the behaviour is described correctly depends upon whether the person ascribing the behaviour has understood the mental state. Thus, in at least some cases, correctly ascribing
the behaviour depends upon correctly ascribing the belief state of the person behaving. If determinations of behaviour depend upon beliefs and beliefs are determined partially on the basis of external conditions, then behaviour is also determined partially on the basis of external conditions.

If behaviour is not supervenient upon the individual, why should we expect the cause of behaviour to be supervenient on the individual? If we grant realism about behaviour, and individualism and supervenience are false of behaviour, then there is no justification for rejecting realism about beliefs because supervenience is violated and individualism is false. In fact, if ascriptions of behaviour are partly determined by contextual factors, then we would be remiss in ignoring those factors when we give an explanation of behaviour via ascriptions of beliefs. Brain states are not, and never could be, acceptable explanations of behaviour for this very reason. While a brain state could explain the movement of Mary's legs, it could never provide an adequate explanation of Mary's driving off to go fishing. If we understand folk psychology as a theory of behaviour, the assumption that behaviour exists is built in. In this case, since supervenience is violated in the case of behaviour it is no longer a problem for belief. It seems to me, then, that if folk psychology is a theory of behaviour, the problem of supervenience is solved and our realism concerning beliefs is as solid as our realism concerning behaviour.
Perhaps, however, rather than demonstrating the validity of the notion of belief, the above sorts of considerations simply demonstrate that our concept of behaviour is in as much trouble as our concept of belief. Both are part of folk psychology and folk psychology is a false theory. If folk psychology is a theory of behaviour it is trying to explain the existence of something that is itself in question. Perhaps we should replace folk psychology with another theory. But folk psychology is supposed to be a theory of behaviour and if behaviour is part of the theory itself, then it is unclear what we are attempting with folk psychology or should be attempting with another theory to explain. We considered bodily movements and determined that although they may escape the context relativity concerns that are at issue with behaviour, they are not really the sorts of things that we are interested in explaining when we use beliefs. But let’s consider them here anyhow.

If we understand folk psychology in more general terms as a theory of bodily movements of humans and animals then the theory would posit both behaviour and beliefs in order to understand the movements of these species. If this is the case though, it is unclear why it would posit behaviour at all, since behaviour is not an explanatory concept but something which we ultimately must also explain. Why posit something just to have to turn around and explain it? If the existence of behaviour were clear, then the positing of beliefs to explain it makes sense. And if
If this were the case, it would also make sense to investigate whether the thing we have posited to explain it really exists. But, as we have just discussed, the existence of behaviour is itself not clear.

It may be suggested, however, that we should not be positing behaviour in the first place and that what we should be interested in explaining is bodily movement as bodily movement, not as behaviour. We should abandon the whole charade of folk psychology, including the concept of behaviour, and focus instead on brain states and bodily movements. This, in fact, is what has been suggested by some eliminative materialists. If behaviour is not something that is physically, individualistically specifiable and beliefs are not physically, individualistically specifiable, one can understand how, within the conceptualization of folk psychology as a theory, the eliminativists get a foothold. But the apparent dilemma of choosing between abandoning folk psychology, which seems practically impossible, and holding onto the existence of non-physically represented entities, is a result of the conception of folk psychology as a theory. If folk psychology is not a theory and doesn’t posit anything, but is instead something like the result of the application of a skill or ability, then the existence of beliefs is not threatened by the fact that they are not physically, individualistically specifiable.
At this point, however, it may be suggested that one need not abandon the view of folk psychology as a theory. Perhaps, it may be suggested, instrumentalism will answer our dilemma. The instrumentalist agrees that we can’t practically dispense with our use of belief ascriptions. There are, as Daniel Dennett points out however, a few varieties of instrumentalism: "Some instrumentalists have endorsed fictionalism, the view that certain theoretical statements are useful falsehoods, and others have maintained that the theoretical claims in question were neither true nor false but mere instruments of calculation" (1987, 72). The position that Dennett himself defends is stronger than either of these. Since we have already discussed some of the difficulties of accepting these weaker versions of instrumentalism in our everyday interactions with one another, we will concern ourselves with Dennett’s variety of instrumentalism. According to Dennett:

...people really do have beliefs and desires on my version of folk psychology, just as they really do have centres of gravity. Do I then grant that the attributions of belief and desires under the adoption of the intentional stance can be true? Yes, but you will misunderstand me unless you grant the following are also true: (1) The gravitational attraction between the earth and the moon is a force that acts between two points: the two bodies' centers of gravity. (2) Hand calculators add, subtract, multiply and divide...It is arguable that each of these is a useful, oversimplifying falsehood; I would rather say that each is a truth one must understand with a grain of salt. (1987, 72,73)

It looks as though Dennett’s form of instrumentalism might allow us to escape our dilemma since it seems that on his view if we
are rational in believing in centres of gravity, we are rational in believing in beliefs. Centres of gravity are theoretical constructs that serve predictive and explanatory purposes just as are beliefs and desires. The concern I raised earlier with regard to our being rational in continuing to ascribe intentional states even though we know there is no basis in physical reality for such ascriptions appears to be answered. Just as one cannot point to something physical and say "there is his centre of gravity", one cannot point to something physical and say "there is his belief that it's raining". This does not make it irrational for us to continue to behave as if these things exist.

There is, however, a distinction to be drawn between beliefs and centres of gravity. When it comes to centres of gravity, we don't expect there to be some physically realized thing that corresponds to a centre of gravity. While the enterprise of searching for some behavioural or neurophysiological equivalent to belief states has occupied philosophers for decades, a similar search for our centres of gravity seems ridiculous. Furthermore, no one is suggesting that we eliminate the theory of gravity because the things it posits, like centres of gravity, don't exist. Dennett might respond that this just proves his point. He may claim that the fact that in one case we are searching while in the other we are not does not indicate a disanalogy. It indicates instead that it is equally ridiculous
for us to be searching for physical equivalents of intentional states.\footnote{As Tom Patton has pointed out to me, centres of gravity are locations which is why a search for physical entities to identify with centres of gravity seems ridiculous. What I argue below is that the only reason that a search for physical equivalents of beliefs does not seem equally ridiculous is because of our understanding of them as theoretical entities. I believe that this way of conceiving of mental states is misguided and that, just like location, a belief, understood in experiential terms, is not a posit in a theory.}

If Dennett were to reply in this way I would be inclined to agree with him. I think, however, that the matter requires further investigation. I think we need to question why philosophers have been searching for physical equivalents of beliefs and desires. And, it seems to me, the answer to this is that the motivation is built right into our conceptual framework of folk psychology wherein the theory is that folk psychology is a theory. If folk psychology is a theory then we should be able to prove it true or false. It looks as though the only way to do this is to either prove that there exist physical equivalents to the states we posit or that there don't exist such physical equivalents. But why is this the case? What is claimed by the theory of folk psychology? Let's consider gravity first.

In order to confirm the theory of gravity, one could predict what would happen in various circumstances according to the theory and then create those circumstances to see if what the theory predicted actually occurs. If it does the theory is
confirmed. If it does not the theory, or at least part of the
time, is proven false. But even if what is predicted by the
time does not occur, we don’t then immediately say that
gravity doesn’t exist. We say that we were wrong about how it
operates. We say this since we know that there is some sort of
force in operation, which we are calling ‘gravity’. We have a
theory about this force and have posited properties of this
force on the basis of what this force would have to be like in
order to explain certain phenomena. Our theory could be wrong,
but the existence or non-existence of some force is not in
question. Or, at least, it is a different question.

A scientific realist may say that gravity does exist while the
anti-realist will claim that it is a useful theory but gravity
doesn’t actually exist. Proving that the theory of gravity is a
true theory is, however, very different from proving that the
theory that gravity exists is a true theory. Both the realist
and anti-realist may accept that the theory of gravity is true
or empirically adequate even though the anti-realist will deny
that gravity actually exists.¹⁸ There is, however, a distinction
between the questions about accepting the truth or empirical
adequacy of the theory and accepting the existence of gravity.

¹⁸ Actually, it seems odd to say that ‘gravity exists’ is a
theory. ‘Gravity explains such and such’ is what is standardly
taken to be the theory and this could be true even if gravity
doesn’t exist.
Now let's consider folk psychology and intentional states. The question at issue differs from questions concerning the content of the theory, the theory of beliefs and desires. We are not here concerned about whether the generalizations of the theory are true. We are not concerned whether, for example, it's true that jealous people make poor partners in relationships or whether, all else being equal, John will open the window when he's warm because he prefers to be cool and he believes that he will become cool by opening the window. If these were our concerns, then there would not be a problem. We could prove and disprove or confirm or disconfirm aspects of our theory of beliefs and desires. Just as in the case with gravity, we could set up test conditions, make predictions and see whether our hypotheses were correct. This, in fact, is just what psychology does. But these sorts of tests do nothing either to prove or disprove the existence of beliefs and desires in the first place. These sorts of tests instead assume the existence of beliefs and desires and test the hypotheses we have about the causal natures of those beliefs and desires. Our concern here is not with the causal nature that the theory attributes to beliefs and desires but with the existence of beliefs and desires as entities. And, according to the view that folk psychology is a theory, beliefs and desires are simply posits in that theory to explain behaviour.
The way to prove that beliefs and desires exist as entities that cause behaviour is obviously to show that they do cause behaviour. But this can't be done by testing any of our generalizations about how possessing certain beliefs and desires will result in certain behaviours because doing so already assumes that the entities exist and that it is their causal properties we are testing. So, for example, if we hypothesize that, all else being equal, pain causes people to wince, we can test this hypothesis by observing people in pain and seeing whether they wince. While this hypothesis does not assume that pain causes people to wince, that is, it doesn't beg the question about what it sets out to prove or disprove, it does assume that pain exists. In the same way, if we hypothesize that Sam left the party early because he was saddened by Jill's rejection of him, we are hypothesizing that Sam is sad, not that there exists an emotion called sadness. This hypothesis is based on another hypothesis which claims that people who are rejected by someone for whom they care deeply are saddened. Again, in this case there is an assumption that something called 'sadness' exists and the hypothesis is that rejection is sufficient to cause this emotion to arise. None of these cases tests the existence of mental states. That they exist is always assumed.

19 A realist might say that, through our tests, we prove that pain causes people to wince, whereas an anti-realist might say that we confirm the theory that pain causes people to wince and that such confirmation counts towards the empirical adequacy of the theory, but not the literal truth of the theory. In either case, however, the prior question of whether there exists an entity 'pain' is not addressed.
The only way, then, that we can test for the existence of the entities posited by folk psychology is to make sure that there are physical equivalents. We need to make sure that there exist physical things that match up somehow with our concepts of beliefs and desires. It seems that the conception of beliefs and desires as posited entities in a theory forces us to start looking for physical equivalents in order to justify our maintaining the existence of the states. And, as we have seen, if externalism is right, our ascriptions can't match up with anything physical because the contents of beliefs are determined partly by social, historical or otherwise non-physical factors.

Eliminativists argue that if the theory of folk psychology is false the entities don't exist. The justification for this appears to be that, other than the confirmation of the theory, there is no sense in which we are justified in claiming that intentional states exist. We have now just examined the fact that it seems that even if we confirm that certain aspects of the theory are true we still do not prove that the entities exist. More seems to be required. It appears to be required that supervenience be true.

With gravity it seems to be the case that we can separate questions of realism about the existence of gravity from questions about the truth or confirmation of the theory. What I've been trying to demonstrate above is that this does not seem
to be the case with intentional states according to the theory view. When we talk about psychological or intentional realism, what is generally at issue is not whether the generalizations of the theory are right, but whether the existence of the posited entities is possible. On the theory view it is not the case that the theory of folk psychology is about the properties of mental states and how they cause behaviour. It is a theory about behaviour and, in attempting to explain behaviour, folk psychology posits mental states. So the existence of mental states is part of the theory. You can’t keep mental states without the theory.

Let’s recall the case of gravity once again. There is some sort of force that is responsible for certain phenomena. We label this force ‘gravity’ and come up with a theory about how it operates. There is no question of looking for physical equivalents of centres of gravity because we understand that they don’t exist in any manner that would result in our finding some physical thing. But since our theory of gravity is confirmed or disconfirmed and modified on the basis of our experiments, we are confident in our views about the theory of gravity. Whether centres of gravity really exist is a separate question whose answer depends on what you require for existence. If you include only things like tables and chairs in your ontology, you would likely be an anti-realist about centres of gravity. But if you included electrons and quarks, then perhaps
you would include centres of gravity. Regardless, however, of your view on centres of gravity as entities, you can be a realist about the theory of gravity.

There are some sorts of forces that are responsible for behaviour. We label these forces ‘intentional states’ and come up with a theory about how they operate. We modify our theory on the basis of our experiments and experience in the world. This, it seems to me, is the approach we ought to be taking in our analyses of folk psychology. But according to the theory theory, it is part of the theory itself that there are some sort of forces responsible for behaviour. Thus, in order to confirm the theory, we must find these forces. Finding these forces means finding something physical. Again, the conception of folk psychology as a theory forces us to search for something physical to justify our intentional ascriptions.

It seems to me that Dennett’s analogy between beliefs and desires and centres of gravity is a good one. And I would agree further that one can understand statements about them as truths ‘with a grain of salt’. I don’t think, however, that this view is sustainable within the theory theory framework since if we grant that it is part of a theory that beliefs and desires exist (rather than that it’s a theory that beliefs and desires have certain causal properties), then the only sense in which this theory can be confirmed is if we are assured of the existence of
some physical things that we can legitimately call beliefs and desires. There is no room in the theory theory for existence 'with a grain of salt'.

But perhaps we are assuming too much if we assume that the claim that folk psychology is a theory includes the claim that beliefs and desires exist. Perhaps folk psychology claims only that particular beliefs and desires cause particular behaviours. And just as in the case with realism vs anti-realism about gravity, we can separate the questions of the confirmation of the theory from realism about the entities. If, however, the claim that folk psychology is a theory is best understood as simply the claim that particular beliefs and desires cause particular behaviours, then, in order to test the theory, the existence of beliefs and desires must be assumed. Thus, in order to know whether the belief that P and the desire that Q cause behaviour X, we must assume that it's possible for people to have the belief that P and the desire that Q. Also we must assume, of course, that behaviour itself exists. And as I discussed earlier, the same arguments that count against the existence of beliefs count against the existence of behaviour, so if we are going to grant the existence of behaviour we are also justified in granting the existence of beliefs.

If this were what was meant by the claim that folk psychology is a theory then, it seems to me, attempts to disprove it would
centre around proving its generalizations false. Furthermore, if
the generalizations did not generally prove to be false, even
anti-realists would not advocate its elimination. The
philosophical discussions, however, are rarely concerned with
the generalizations of folk psychology. They are concerned with
the possibility of physical realizations of intentional states.
They are concerned with the existence of these entities posited
to explain behaviour. And the arguments against folk psychology
have been arguments against the possibility of reduction.

The position that is advanced by Dennett is a position with
which I, for the most part, am in agreement. The problem is that
it is not a position that is possible within the conceptual
framework of folk psychology as a theory. Dennett's position has
been understood as instrumentalism because he doesn't believe in
supervenience, in entities physically equivalent or identifiable
with intentional states, but he does believe in the
generalizations of folk psychology. I have argued that on the
theory theory construal, the theory of folk psychology is not
about the generalizations of folk psychology, but rather it is
a theory of behaviour in which mental states are posited. As a
theory of behaviour in general, the theory is that intentional
states cause behaviour. It is part of the theory of folk
psychology that intentional states exist. Thus, we cannot accept
the theory without accepting the existence of the entities. And
we must, if we are to avoid begging the question, have some
independent ground from which to justify our claim that the entities exist. We cannot simply point to successes in the predictions of the theory. Dennett’s view is not an option within this scenario.

If you accept the theory, as Dennett does, you must accept the existence of mental states, and if you accept the existence of mental states, you must justify your position through physicalism. Thus supervenience must be true. But Dennett denies supervenience. Within the framework that conceives of folk psychology as a theory Dennett’s view appears to be inconsistent.

Within an alternative conception of folk psychology, however, Dennett’s view makes a great deal of sense. On an alternative conception we have beliefs just like we have centres of gravity and just like calculators add, subtract and multiply. The question of whether we really do have beliefs becomes as significant as whether centres of gravity exist or calculators really do add. Just as we wouldn’t suggest throwing out our calculators because we can’t find signs of intelligence within, we wouldn’t consider eliminating folk psychology because we can’t find physical equivalents of beliefs within. But this is true only if the existence of beliefs is not a theory that we are trying to prove.
An alternative conception of folk psychology that would allow us to maintain realism, then, would still be able to maintain that folk psychology is a theory in the sense that the generalizations it makes are empirically testable. Thus when I say of Jack that he went home early because he wanted to avoid the traffic, I am presenting an hypothesis about the cause of Jack's behaviour. The hypothesis suggested by the theory must be that Jack's behaviour was caused by mental state A or B or C, for example. The hypothesis suggested by the theory is not that Jack's behaviour was caused by A or B or C, rather than brain states or acts of God. That beliefs cause behaviour is not what gets tested. This must be granted in order for the theory even to get off the ground. Which beliefs cause which behaviours are the theoretical postulates.

The theory theory claims that when we ascribe beliefs we are applying a theory. We posit beliefs to explain behaviour. This is true in the sense that we posit one particular belief rather than another but not in the sense that we posit beliefs at all. It seems to me that we can't help but see people as having beliefs. Neither can we help but see people as engaging in behaviour. Conceptually, there are not two steps involved where we first observe behaviour and hypothesize that it was caused by a mental state and then next hypothesize which mental state. But then I'm not applying a theory when I ascribe a belief to Jack; I am only using a theory to decide which belief.
Is folk psychology a theory then? Not in the way it has been standardly understood. It has been standardly understood to be the case that folk psychology is a theory of behaviour and that, very crudely and simply put, the theory is that beliefs and desires cause behaviour. If I am right, however, and we are not applying a theory when we explain behaviour (we may be using a theory to decide between explanations) then how do we come up with the idea that behaviour is caused by beliefs? If it's not a theory that behaviour is caused by beliefs, what is it?

The alternative theory of folk psychology which is currently receiving attention is called the 'simulation theory'. According to the simulation theory, our explanations of behaviour via intentional states are the result of a type of cognitive capacity or skill rather than a type of theorizing. If this type of view is correct, then perhaps when we ascribe intentional states to others we are not actually positing beliefs to explain behaviour; we are simply somehow cognitively perceiving what they do. According to the simulation theory:

...human beings are able to predict and explain each others' actions by using the resources of their own minds to simulate the psychological aetiology of the actions of others. So, instead of being theorizers, we are simulators. We are mental simulators, not in the sense that we merely simulate mentation, but in the sense that we understand others by using our own mentation in a process of simulation. (Davies and Stone 1995b, 3)
The focus of the literature on the simulation theory of folk psychology vs the theory theory has not been on the relationship between propositional attitude realism or the integrity of psychology and how simulation theory may help save it. Rather, the focus has been upon explaining certain phenomena in developmental psychology. Proponents of each view present their cases in light of how their theory better explains the psychological phenomena. In addition to examining how children appear to gain quite suddenly the ability to ascribe beliefs appropriately to others, a great deal of attention is being paid to studies on children with autism and how poorly they do on theory of mind type tasks in comparison to control groups with similar or lower levels of traditional intelligence.

Although those participating in the debate usually claim that the matter is an empirical one, it seems that at least so far, much of the data can be equally well explained by either theory. The simulation theory, however, is the more recent proposal and as such it has fared very well in presenting a challenge to the traditional view. Thus, if externalism is true, as is our assumption, the demise of psychology as we know it does not necessarily follow. If there is an alternative way of conceptualizing our ascriptions of mental states, perhaps there is hope for integrity in psychology after all.
That the simulation theory does pose an alternative to the eliminative materialist route is evident. Stephen Stich and Shaun Nichols, who examine and critique the simulation theory as it is presented by Robert Gordon and Alvin Goldman, explain it this way:

The central premise in the eliminativist's argument is that neuroscience (or connectionism or cognitive science) is on the verge of demonstrating persuasively that folk psychology is false. But if Goldman and Gordon are right, they will have pulled the rug out from under the eliminativists. For if what underlies our ordinary explanatory practice is not a theory at all, then obviously it cannot be a radically false theory....Gordon and Goldman claim that the theory which posits a tacitly known folk psychology is itself radically false, since there are much better ways of explaining people's abilities to interpret and predict behaviour. (1995b, 124-25)

While it may be clear that if folk psychology is not a theory it can't be a radically false theory, we must demonstrate, on the positive side, just how simulation theory can restore our sense of justification in believing in beliefs. If this alternative conception is correct, I understand your behaviour not by positing beliefs in you, but by, in a manner of speaking, putting myself in your position. Beliefs are still explanations of behaviour, but only because I can understand your behaviour by experiencing the same context-bound beliefs.

In order for the case to be made that folk psychology is justified within the simulation theory conception, we must gain
a better understanding of what is being proposed by simulation theory. I would like first, however, to review the requirements of an alternative theory in order for it to be the case that psychology is justified. We have identified several problems within the theory theory framework and it is these problems to which we must find a resolution within simulation theory.

We began with the problem of supervenience that arises as a result of the externalist arguments. We concluded that the supervenience problem could be solved within the theory theory but that the solution was not satisfactory. The solution depended upon the analysis of behaviour which showed that behaviour itself is not supervenient. The fact that behaviour is also not individualistically realized will lead to realism about beliefs as a conclusion only if we assume, despite its context-bound nature, realism about behaviour. But if we are questioning the existence of beliefs on the basis the fact that they are not supervenient, we are not justified in assuming the existence of behaviour. Furthermore, in addition to the fact that behaviour violates supervenience, there doesn't seem to be any principled way to distinguish cause-associated beliefs from behaviour-associated beliefs. Whether Jane's behaviour when she is running down the street is 'going for a jog' behaviour or 'running from a perceived threat' behaviour depends on the associated beliefs. To view someone as behaving, rather than just moving, we have already attributed beliefs. If one is offering a theory of
behaviour, offering beliefs as the cause of behaviour, it's not clear how that person decides what beliefs count as part of the description of behaviour and what beliefs would count as the cause. It's not clear how one knows what beliefs to posit to explain behaviour. If we decide to say of Jane that she is going for a jog, we have already attributed to her an intentional state. We can then ask for an explanation of this behaviour which involves a further intentional state. She may be jogging because, for example, she wants to get in shape or because she wanted some fresh air, or both. Although in practice we seem to be able to separate our descriptions of behaviour from our explanations of behaviour, there appears to be no principled way of doing this. Theoretically, separating the explanandum from the explanation appears to be a problem.

From the simulation perspective the problem of the failure of supervenience threatening realism is solved. The solution offered above, i.e. that we experience the existence of beliefs, is satisfactory. It is satisfactory because the existence of beliefs is never challenged on the basis of their context-bound nature in the first place, and the fact that we don't know on what basis we are able to distinguish behaviour-associated beliefs from cause-associated beliefs is irrelevant. The fact that we do make these distinctions and that we understand one another on that basis is what is relevant. What is relevant, in other words, is that by simulating your states, by putting
myself in your place so to speak, I am able to know what you think, not by applying any theory, but because we are both human beings who share similar physiologies, and, as a result, we react to the world in similar ways. The simulation theory proposes that we understand one another not by positing mental states, but by somehow 'projecting' ourselves into the situation of the other, and in a manner referred to as 'off-line', because we are imagining the context rather than actually being in it, experiencing the same states. In this case we don't have to know the conditions of ascription of mental states, we simply 'experience' those states.

Finally, the problem of whether there is any truth to the aspect of our theory that says that beliefs exist actually disappears since it is not considered theoretical that beliefs exist. They simply are our means of explaining and predicting behaviour.

The externalist arguments are accommodated on this view. The question of psychological realism can now, it seems to me, be appropriately lumped together, in the way that Dennett suggests, with questions about realism of other things like centres of gravity. There is nothing to distinguish beliefs from these other sorts of abstracta. Thus Dennett could finally escape the instrumentalist label and legitimately call himself a realist. Burge, who considered himself a realist all along, could also
comfortably maintain his position within the conception of folk psychology as described by the simulation theory.

In this chapter I have argued that externalist claims considered within the framework of the theory theory force us to abandon realism about folk psychological concepts, that we can't practically dispense of these concepts and that, analogously to the case with induction, we shouldn't even be seeking to demonstrate their existence. If I am right in my arguments so far, and if the simulation theory can allow us to maintain realism about the concepts, the simulation theory receives support on independent non-empirical grounds. Alternatively, and more to the point in this chapter, if the simulation theory is a plausible alternative to the theory theory and it can accommodate externalist arguments, then realism and integrity in psychology need not be abandoned. How the simulation theory presents a plausible alternative is the focus of the next chapter.
Chapter Four

Simulation and the Context Bound Nature of Belief Attribution

The simulation theory claims that folk psychology is not a theory. When we ascribe beliefs and desires to others we do not do so on the basis of a theory. Folk psychology is still the ascribing of intentional states to explain and predict behaviour. But when we ascribe intentional states we do so in a manner that is not theoretical. According to the simulation theory, rather than forming hypotheses by processing information, we project ourselves into the other's circumstance.\(^{20}\)

\[^{20}\text{The term 'projection' refers to placing one's self in the situation of the other. When this is done, one is not simply imagining herself as being in the other's shoes, so to speak, but is imagining herself as being the other. It is different from the explicit exercise of imagining what you would do if you were in her shoes.}\]

The externalist arguments point out that the contents of psychological states can vary while the brain states remain the same. The contents of one's psychological states are at least partially determined by external or environmental factors.

Imagine, for example, that your friend Mike just won a million dollars. When projecting you wouldn't imagine that you won a million dollars, but rather that you were Mike and you won a million dollars. Within this projection you 'experience' the beliefs that Mike would experience except that since you are not actually in the circumstance, since it is an imagined or 'simulated' world in which you are Mike winning a million dollars, the experiences you have are referred to as 'off line'. They are on a different track than non-projection based beliefs.
Supervenience is violated and, as a result, psychological realism is claimed to be impossible. Like Burge and Dennett, I believe that realism is possible without supervenience. Unlike Burge and Dennett, however, I believe that this option is only available through the rejection of the notion that folk psychology is a theory.

Since my ultimate concern is to defend realism, and I am claiming that simulation theory enables us to do so, we need to consider the particular externalist arguments against realism within the context of simulation theory. First, however, some further clarification of the simulation theory is in order. To this end I will briefly review a couple of the studies that inspired the idea of this alternative framework. I will then show that on simulation theory externalist concerns are answered and that the explanations we use that invoke intentional states are causally relevant.

Although the issue of concern for those conducting the studies is explaining developmental differences rather than realism in psychology, the studies do illustrate the original motivation behind the simulation theory and shed some further light on the subject for our purposes. The ‘Sally-Anne Experiment’ conducted by Simon Baron-Cohen, Alan Leslie and Uta Frith, experts in the study of autism, is a false belief task in which children with autism, normal children and children with Down’s syndrome, all
with a mental age of above three, were asked to identify a belief which is false:

We used two dolls. Sally and Anne, and acted out a little scenario: Sally has a basket and Anne has a box. Sally has a marble and she puts it in her basket. She then goes out. Anne takes out Sally’s marble and puts it in her box while Sally is away. Now Sally comes back and wants to play with her marble. At this point we ask the crucial question: ‘Where will Sally look for her marble?’...The failure of the autistic children to understand Sally’s belief is...remarkable, because they actually had a much higher mental age than the other children. Intellectually they were able to solve a great many logical problems. (Frith, U. 1989, 160-61)

The children with autism fail to attribute to Sally the false belief that the marble is in the basket. Since Anne put the marble in the box, the children with autism say that Sally will look in the box, even though Sally had no way of knowing that the marble was in the box. They don’t seem to understand that Sally’s state of mind is relevant to where she will look. This study was inspired by an earlier one by Heinz Wimmer and Josef Perner that noted the developmental differences between normal children at age three and older children. The three year olds failed to identify the false belief:

...children were shown a puppet character, Maxi, who put his chocolate in a box and went out to play. Next they saw Maxi’s mother transfer the chocolate to the cupboard while Maxi was still out. Where will Maxi look for the chocolate when he returns? Five year olds responded: in the box; a response typical of adults. Children of three, however, indicated the cupboard, where they themselves believed it to be. So three year olds have not fully mastered the art of belief ascription.
This three year old phenomenon is quite robust, and applies even to their own previous false beliefs (Astington and Gopnik, 1988; Wimmer and Hartl, 1991). Shown a closed candy box, children express the belief that there are candies inside. The box is open and seen to contain only pencils. The children are then asked what they had originally thought was inside. Three year olds typical insist that they thought there were pencils, whereas four and five year olds admit that they thought there was candy. (Goldman, Alvin 1995a, 193)

The studies can be interpreted as indicating that children under three and children with autism have not developed a theory of mind. That is, they lack a theory that says that behaviour is explained by beliefs and that may even include particular examples such as that, unless one knows that an object has been moved by someone or something else, one will be caused to look for the object where one last saw it by the belief that that's where it will be. The studies could also be interpreted as indicating that these children do not lack an ability to theorize about mental states, but some other sort of skill or ability. Since the children with autism are not lacking other intellectual skills and are able to theorize about other things, those studies in particular appear to reflect favourably on the second interpretation. Another point that has been raised in defense of the second interpretation is the fact that children with autism also lack an ability to pretend. In fact, a lack of pretend play is one the more significant behavioural deficits to note in diagnosing a child with autism. Since simulation is a type of imagining or pretending, this deficit would be consistent with what simulation theory would predict:
...our access to the thought of others is not through the application of a primitive but effective theory, as advocates of the 'theory-theory' of folk psychology suppose, but through a kind of internal, largely spontaneous, re-enactment that allows us to imagine ourselves in some rough approximation to the situation of another. In so imagining, we tend to acquire, in imagination, the beliefs and desires an agent would most likely have in that situation, and those imaginary beliefs and desires have consequences in the shape of further pretend beliefs and desires as well as pretend decisions that mimic the beliefs, desires and decisions that would follow in the real case. And all of this happens as an initial causal consequence of the initial imagined state, and not by the application of any folk theory of mind, which simulationists think would have to be so complicated that it could not plausibly be credited to any but the most sophisticated among us. (Gregory Currie 1995, 158)

Given this new framework, where imagination rather than theory does the work, let's now consider the externalist claims. According to the externalist arguments, what X believes is partially determined by the context in which he is believing. If what X believes is partially determined by external conditions, then when I simulate X to determine what he believes, the off-line beliefs I have are also subject to external conditions.

If I simulate Daniel who goes to the doctor I will conclude that he has arthritis in his thigh and my off-line belief will be false in the actual case and true in the counterfactual case since I am subject to the same contextual determinations as Daniel. In the Twin Earth case, if I simulate Oscar2 and attribute to him the thought that 'water is wet', then my off-
line belief will be subject to the same contextual conditions and 'water' will refer to twater (XYZ rather than H\textsubscript{2}O).

But perhaps simulation theory goes further than is required. If I determine what you believe through simulation rather than through the application of a theory then, it may be claimed, it is not clear how my claim about what you believe is in any sense an hypothesis. Thus it is not clear in what sense I may be able to be wrong in my ascriptions to you. The response to this is that regardless of how I come to ascribe the belief that P to you, it is still a type of hypothesis that you actually hold the belief that P, in that it is a statement that can be shown to be false. What isn't an hypothesis is that I am correct in ascribing an intentional state to you, i.e., that some intentional state or another is the cause of your behaviour. I do not attribute an intentional state as an explanation of your behaviour on the basis of a theory. It's just something I do.

The following analogy may help to clarify both the nature of the status of an attribution of an intentional state and the way in which such attributions can be wrong: My car won't start. I believe that my car won't start because I haven't put oil in it for six months and, since I am the only one driving the car, I have privileged access to this information. In order to come up with his hypothesis, my mechanic looks under the hood. He comes to the hypothesis that I have been neglectful about putting oil
in my car. Although we used different methods, we both reached the same hypothesis about why my car won’t start. According to the simulation theory, my access to the hypothesis that John’s behaviour is caused by the belief that P is much more like my access to the hypothesis that my car won’t start because I didn’t put oil in it than it is like my mechanic’s access to the hypothesis that my car won’t start because I didn’t put oil in it. It may in fact have turned out to be false that that was why my car didn’t start and it may in fact turn out to be false that John behaved as he did because he believed that P. But the way I knew about the oil in my car is the same way that I know that the belief that P fits the circumstances I have projected myself into. This type of knowledge is not based upon theory. Theory is not necessary for me to surmise that I didn’t put oil in my car and it is not necessary for me to surmise that if John were in the context I projected he would believe that P.

It is, however, an hypothesis that the car won’t run because I forgot to put oil in it and it is an hypothesis that John engaged in behavior X because he believed that P. What happens then when these hypotheses turn out to be wrong? It might have been the case that although I forgot to put oil in the car, that there was enough oil there already so that no damage resulted. It may have been that something was wrong with my starter mechanism and that was why my car wouldn’t start. Similarly, although it would have been the case that John would have
believed that P, given the circumstances I projected myself into, my projection neglected some important elements of John’s personal history so that the belief that P would not cause behaviour X in John’s case.

For example, assume behaviour X is ‘quickly leaving the room’. According to simulation theory when I observe and attempt to explain this behaviour I do so by projecting myself into John’s position and simulating John’s mental state, by experiencing ‘off-line’ what he is actually experiencing. So, for example, I may explain his behaviour by saying that he didn’t like the company and was bored with the conversation. It may turn out to be the case, however, that John is allergic to cats and a cat had just entered the room. I may, thus, fail to project myself fully and to fully represent his context. My claim that John engaged in behaviour X because he believed that P is thus an hypothesis but the claim that if John’s context had been as I represented it to be, he would have believed that P is not an hypothesis. Thus cases of failure in attributions are not failures in theoretical knowledge; they are failures in imagination.

The difference between the theory theory and the simulation theory, then, is in their claims of how we arrive at the hypothesis that John’s behaviour X was caused by the belief that P. In addition, however, according to the simulation theory,
it's not just that I explain John's behaviour in this instance as being caused by the belief that P because when I simulate John that is what I believe, but the fact that the belief that P can cause behaviour X at all is something that I just know through experiencing the causal properties of the belief that P. When I simulate John and experience P, I experience it as the cause of John's behaviour. It is the manner through which I understand John's behaviour. Here there is a disanalogy with the car example. Although I have privileged access to the fact that there isn't much oil in my car, I had to learn, as part of a theory, that lack of oil may cause damage to a car.

That the belief that P can cause behaviour X is something that I do not learn through theory. When, during simulation, I experience the belief that P it is not experienced as distinct from its causal properties. The belief that P is a causal entity and is experienced as such. For the theory theory it is theoretical that: 1) There is an entity that is the belief that P; 2) The belief that P can cause behaviour X; and 3) The belief that P caused John's behaviour X. For the simulation theory it is not theoretical that there is an entity that is the belief that P nor that the belief that P can cause behaviour X. The only thing that may be considered theoretical, in that it is a falsifiable statement based on the evidence of simulation, is that the belief that P caused John's behaviour X.
Let’s now return to our Twin Earth example and consider how error may fit into the picture. Recall that I claimed that in the Twin Earth case, if I simulate Oscar2 and attribute to him the thought that ‘water is wet’, then my off-line belief will be subject to the same contextual conditions and ‘water’ will refer to twater (XYZ rather than H₂O). I may, however, fail to place myself properly in his context and attribute the wrong belief. If my simulation doesn’t include the fact that Oscar2 is on Twin Earth, then when I attribute the belief that ‘water is wet’, ‘water’ refers to water and not to twater. In this case, because of a lack of fully simulating, I will attribute the wrong belief. But this is not a problem. We do sometimes attribute wrong beliefs and on this view it is because we are not aware of all of the conditions affecting the determination of belief.

The beauty of the simulation view is that in most cases not knowing the conditions will not, however, affect our ability to attribute the correct belief. Since most of the time when we are interacting with people we are interacting with people for whom the conditions for determination of particular beliefs will be shared, it is irrelevant whether we are actually aware of those conditions. That is, we interact with people for whom water is H₂O and arthritis is an inflammation of the joints. For our present purposes, it doesn’t really matter what the precise conditions of belief determination are. What matters is that when I simulate you, I am subjecting my off-line beliefs to
those same conditions to the extent that I have placed myself in the same context.

If when I simulate you, I fail to place myself completely in your context, then to the extent that the areas in which I fail to simulate the context are important in belief determination, I will fail in my attribution. In fact, I will never be able to simulate you fully or fully place myself in your context because your context will always include individual private facts of which I am unaware. Whether those individual facts that form part of your context have any bearing on meaning determinations is not a question we need to pursue at this point. The relevant point is, even if they do have some significance, the significance of those facts is apparently typically outweighed by the social, shared external facts, which is why we are, for the most part, successful at understanding one another. What is important is that given the same context, we will behave in the same ways and on the basis of the same beliefs. And the manner in which the context determines your belief will be the same manner in which the context determines my belief.  

Notice that this will remain true even on an individualistic type of theory of meaning. If, for example, when you were taught the word 'cat', what you were actually pointing at and labelling was a pig, the manner in which the environment contributes to the meaning of 'cat' for you would still be the same as it would be for me if I am simulating you, placing myself in your context. Of course, in such a case, since most people do not learn the meaning of 'cat' in such a manner, this is not a shared external fact and I will likely fail at my simulation.
The contents of my belief will be determined, just like yours, in part on the basis of external factors which don't always get physically represented. Since, however, I understand and explain your behaviour on the basis of beliefs that I am myself experiencing off-line, the fact that external factors aren't always physically represented does not affect my ability to ascribe beliefs to you. In fact, it may even be the case that your beliefs are not fully determined. If I am simulating you, my beliefs will also not be fully determined. The criteria for the contents of beliefs to serve as an explanation of behaviour has thus changed from being physically represented to being the belief contents that I would experience given your context which includes both internal and external facts.

If I have placed myself in your context then I will experience (off-line) what you experience. If you are on Earth, it's raining and your roof is leaking, I will explain your bucket-fetching behaviour by experiencing your situation and surmising that you don't want water damage in your house. If you are on Twin Earth, a full simulation will include this fact and I will

---

22 On Daniel Dennett's view, for example, meaning and intentional state content are not always fully determined. On Dennett's view, the external factors that determine content are teleological, evolutionary factors. He talks about 'mother nature' determining content. But, in some cases, it is not clear that mother nature intended one interpretation rather than another. When two interpretations, or two assignments of content, are equally useful there may actually be no answer to what something really means. One interpretation may be as right as another as long as both have equal utility.
explain that you don't want water damage in your house. If I fail to project myself fully into your situation, the explanation that you don't want water damage is incorrect. If, for example, I observed your behaviour on Twin Earth and told you of my explanation you might say that since you don't know what 'water' is you are not concerned about that. What you are concerned about rather, is water damage. The fact that the explanations cannot be substituted illustrates that although content might not supervene upon brain states, it is relevant to explanations of behaviour.

If I do not simulate your context fully, there is always the danger that the areas in which I fail to simulate will be significant to the determination of the belief content. But, as I mentioned earlier, it seems to be most often the case that the areas in which we fail to simulate fully, such as personal individual histories, do not affect the determination of the belief content. Of course, in many cases such histories do make a difference and we must make a conscious effort to understand the behaviour of someone. When this happens our simulation becomes explicit and we bring to bear more pieces of information than those that were readily available to us initially. In these types of cases, when we have difficulty understanding someone's behaviour, it is because our built in assumptions have somehow failed. The young child crying at the supermarket is handed a cookie and you fully expect the child to stop crying but he
doesn’t. You must then actually think about what is causing him to cry. You notice that it’s very warm and you change your ascription from the desire for a treat to the desire to be cooled down.

But this isn’t the type of case that is cause for concern. We are concerned with those cases in which your ascription fails not because of anything that is physically represented as distinct, as crying for a cookie would surely be distinct from crying because one is overheated, but because the belief you ascribe is distinct from the one being held purely on the basis of external, environmental conditions. The existence of such cases was what was demonstrated by the examples of Burge and Putnam. The question is whether these environmentally distinct contents can be distinguished in terms of causal efficacy. I am arguing that since my explanation of your behaviour will be incorrect unless I properly simulate your context that these environmental conditions are in fact causally significant. In context E, the cause of your behaviour was that you didn’t want water damage. In context TE, the cause of your behaviour was that you didn’t want twater damage.

‘The supervenience problem’ is that we must abandon realism over beliefs since in order for beliefs to have causal efficacy they must supervene on brain states and Twin Earth examples of beliefs concerning water are examples where belief contents
don't supervene on brain states. I am claiming that the twin examples are also examples of situations where the belief contents that don’t supervene on brain states do have causal efficacy. That they have this efficacy is demonstrated by the exercise of simulation.

If I fully simulate John(te), my beliefs will be about twater. If I don’t and I ascribe to him a belief about water I will be wrong. Although John himself doesn’t know the difference between water and twater, he behaves and believes in accordance with the implicit assumption that when he thinks about water he is thinking about the stuff that’s called ‘water’ on his planet or perhaps, by his community, or by the experts on his planet. Whatever the precise nature of this implicit assumption, what is clear is that it would rule out a substance that nobody on his planet refers to when they use the term ‘water’. Although we may not know how to define John’s linguistic community, we know that it won’t include Earthlings just as our community won’t include Twin Earthlings. Of course, if Earthlings also lived on Twin Earth this might change. But then the external conditions determining content will also have changed and when I simulate John, this aspect of his context must be represented in my simulation if I am to ascribe the correct belief.

John(te) would not have grabbed a bucket if his belief were that he didn’t want water damage. Since the implicit assumption about
'water' is different from the implicit assumption about 'twater' having such a belief would be strange indeed for John(te). Living on planet Twin Earth, John(te) simply would not hold the implicit assumption about 'water'. Even if John(te) and his twin had been told in advance that water is pretty much the same thing as twater, so that when they saw rain dripping through the roof they were still in the same brain states, it would not be possible for him to have a water belief when he sees the rain, since his environment determines that 'water' refers to XYZ. The only way for him to have 'water' H$_2$O beliefs is if he consciously makes the distinction. If, for example, he saw the rain and thought, "I wonder if that is water rather than twater". Once he does this, his brain state changes. In order for John(te) to think, "I don't want water damage", rather than "I don't want twater damage", he would have to have belong to a different linguistic community, an Earthbound community. This is true unless the thought were simply somehow planted in him rather than caused by seeing the rain. But then, upon having the thought, he would not be caused to go grab a bucket. Rather he would be caused to stop and wonder why he would think about a substance with which he has had no experience. Or perhaps he would be caused to stop and wonder whether he was going crazy.

There is agreement that the thought that 'water is dripping' is a different thought from the thought that 'twater is dripping'. Call the first thought 'A' and the second thought 'B'. What I am
claiming is that, short of divine intervention, it is not possible for John(te) to have thought A since neither he nor anyone else in his linguistic community he has ever heard of or encountered the type of thing referred to in thought A. Furthermore, if John(te) did have thought A, it would not explain his behaviour. Thought B is an appropriate explanation of John's behaviour, but the thought that H2O, the substance that 'water' refers to within another linguistic community, is dripping from the roof, does not, in itself, explain why John(te) went to grab a bucket. John(te) would have behaved differently if the content of his belief referred to something other than what it does. If I thought that a substance from another planet was dripping from my roof I would run out of the house rather than worry about water damage.  

Let's consider Burge's arthritis example in light of this analysis. If I didn't know that arthritis was an inflammation of the joints and I had a pain in my thigh which I thought was arthritis, I would be alarmed. Just like in Burge's example, I would go visit the doctor. If, however, the content of my belief referred to something other than what it does, I would behave differently. If I thought that a substance from another planet was dripping from my roof I would run out of the house rather than worry about water damage.  

---

23 One way in which it would be possible for John(te) to have a water thought is if he had suddenly, during his sleep, been somehow transported to Earth. He would have no idea he is on another planet. Imagine that he looks at the roof from which water is dripping. In this case what he is thinking must be that 'there is water dripping from the roof'. He must be thinking a 'water' thought rather than a 'twater' thought because the external environment determines the content of his thought. The only way he could be think a twater thought on Earth is if he explicitly distinguished between water and twater and Earth and Twin Earth. In this case, however, the underlying brain states are also distinguishable. 
was about tharthritis, a general rheumatoid condition, I would likely not be so concerned, and I probably would not go see a doctor. I have modified Burge's example here to illustrate my point. My behaviour differing in the counterfactual situation depends upon my holding the implicit belief that one should only visit doctors for serious conditions and that general rheumatoid ailments are not serious conditions. This implicit belief could be held by me in both the actual and counterfactual cases so that our brain states remain the same. Given these conditions, however, while the belief that I have arthritis in my thigh explains my behaviour in the actual case the belief that I have tharthritis in my thigh does not explain my behaviour in the counterfactual case. My brain state remains the same, I believe that I have 'X' in my thigh, but because the external conditions differ, the arthritis explanation doesn't work in the counterfactual case. I wouldn't go see the doctor about a general rheumatoid ailment. In the counterfactual situation the more appropriate explanation would be something like that I thought that I had some kind of serious condition in my thigh. But if 'arthritis' refers to a general rheumatoid condition, it would not explain my behaviour.

We have thus far considered three points in favour of external conditions having causal efficacy even though they don't supervene on brain states. Firstly, when I simulate you the external conditions that determine the content of your belief
also determine the content of mine and, therefore, it is not possible for an alternative belief, which is determined by different external conditions, to serve as an explanation of your behaviour. On the simulation model, the external conditions are built right into the explanation. Secondly, it's not just that an alternative belief can't serve as an explanation because we are both constrained by the same external conditions. Because of the implicit assumption that the contents of the concepts we use are determined in the same way for us as for the rest of our linguistic community, it is not possible for someone in the same context to have a belief with an alternative content. Thirdly, even if there were some way in which we could ascribe the alternative or twin belief to someone, when we imagine this situation, the resulting behaviour would be different. So, even if we assume the same brain states and we somehow plant this other referring belief in Earth John, it would be unlikely that this alternative belief would cause the same behaviour. If John had a belief about twater dripping from his roof he would run rather than grab a bucket and if Mary had a belief about tharthritis in her thigh she would take an aspirin rather than seek medical help since when we ascribe beliefs the content is determined by our context. If I ascribe to John the thought that 'water is wet', I must be ascribing a thought about H₂O, unless I consciously make the distinction and then ascribe to John the thought that 'twater is wet'. Thus, if I ascribe 'twater is wet'
to John this won’t explain his drinking it. If indeed this were his thought he would behave differently. I wouldn’t drink XYZ.  

There is one more point to be considered, which draws on our discussion in the last chapter on the context-bound nature of behaviour. Recall that we concluded that the concept of ‘behaviour’ was in the same trouble as the concept of ‘belief’. If there is no sense in which we can understand some movement as ‘behaviour’ unless we view it in the context of the environment in which it occurs and in the context of being performed by an intentional system, then the environmental conditions that partially determine the behaviour will be the same environmental conditions that partially determine the belief. If we ascribe a different belief, then the behaviour may also be reinterpreted. Thus, on Twin Earth we would understand John(te)’s behaviour as

---

24 If our explanations are causally significant, as realism requires, we shouldn’t be able to substitute the belief that P for the belief that Q where P is a water thought and Q is a twater thought. Given the violation of supervenience, there is nothing to prevent this substitution on the theory theory but, on the simulation theory, it is not possible to substitute because we ascribe beliefs in a context.

So simulation theory will allow us to count different explanations as causally different because they will individuate different underlying causes in different contexts. This is not to say that differences in explanations will always count as causally significant. There may be two different, context-bound ways in which to individuate the same cause. Whether two different explanations individuate the same cause is determined by external, contextual conditions. Just how this determination is achieved is not the problem of concern in this thesis. Explaining, in those cases where we agree that there is a difference—such as in the case of water and twater—how that difference can be causally significant when supervenience is violated is the problem.
'fetching a pail to avoid twater damage behaviour', whereas on Earth we would understand it as 'fetching a pail to avoid water damage behaviour'. Since there is no principle for deciding how to distinguish what aspects of the environment and intentions of the person behaving are part of the behaviour and what aspects are part of the belief, there is no in principle reason why we cannot interpret the behaviour in this way. Therefore, since we are explaining distinct behaviours, despite the fact that the brain states are the same, our distinguishing the causes makes sense. It seems to me that if we are explaining distinct behaviours, it actually makes more sense to distinguish the causes than to avoid violating supervenience. Supervenience is violated at both ends.

If the contents of beliefs can be causally efficacious, insofar as a change in content would result in a change in behaviour even when the contents are not physically represented, then we have shown that supervenience is not required for realism about beliefs. I believe that the above considerations demonstrate this conclusion. In order to accept the above considerations, however, one must also accept that our method of understanding each other is not through the application of a theory but is through simulation. If this is not accepted then the first three points may be rejected and the point about the context-bound nature of behaviour will lead, as we previously have discussed, to the rejection of realism about beliefs.
If we understand one another on the basis of a theory rather than through simulation, then the first point about my simulated beliefs being constrained or determined by the same conditions as yours is irrelevant. If the theory theory is correct then although the contents of our beliefs may be determined by the same conditions, in order to explain behaviour we must understand those conditions and abstract from them. On the theory theory view having a belief or experiencing the same contents is not an explanation of behaviour. One must understand and explain how the belief causes behaviour.

If we posit beliefs to explain behaviour there is no built in assumption that what we are positing is constrained by external conditions. We make claims such as that the belief that 'water is dripping from the roof' will cause behaviour X. It can do this because it is physically represented. Thus it becomes a problem and the theory is disproved when the belief that 'water is dripping from the roof' can explain behaviour X because it would require a different physical representation than on the theory theory view. Since we have posited some objective thing to explain behaviour, there can be no causation without physical representation. Within the simulation theory the explanations can't be substituted. Although the brain states remain the same, the causes of behaviour are different since the cause is not some objective thing posited by a theory, but something discovered through experience. The criterion has changed from
not violating supervenience to not violating experience. If we experience different beliefs through simulation, then regardless of whether those beliefs supervene on the same brain state, they are distinct beliefs as well as legitimate explanations.

Our second point regarding the implicit assumption that the contents of the concepts we use are determined in the same way for us as for the rest of our community would not necessarily be rejected by the theory theorist. The implicit assumption point is that it's not possible to ascribe twin contents because the implicit assumption would be violated. The implicit assumption about water, recall, is that when we think about water we are thinking about the stuff that's called 'water' on this planet or perhaps, by this community, or by the experts on this planet. Whatever the precise nature of this implicit assumption, what is clear is that it would rule out a substance that nobody on this planet refers to when they use the term 'water', like XYZ. Thus ascribing the belief to Earth Mary, 'I have tharthritis in my thigh' would violate her implicit assumption. We couldn't equally well explain Mary's behaviour with this belief since it's not possible for Mary to have this belief. When she thinks about arthritis she is thinking about what her community means by 'arthritis'. The implicit assumption is compatible with a simulationist theory, but it is something that a theory theorist would need to argue for more explicitly.
If we were to construct a theory of explanation of behaviour, what grounds would we have for assuming that the meaning of what I say has the same determinants as the meaning of what you say when we utter the same sentence? And how would we decide when the environment has changed enough so that now the determinants of meaning are different? If we posit beliefs on the basis of a theory, we need to know when to change the content of the beliefs we are positing based on the environment. If we explain behaviour by experiencing beliefs in context, we don’t need to know any theory of meaning. There just has to be one. Thus, in order for the theory theorist to utilize this point in defense of realism, she would need to show exactly how contents are determined. That is, she would need to show whether contents are determined by experts, or one’s country, or the human race, or the physics of the universe, etc. Furthermore, if we are to be correct in our ascriptions to others, we would all have to know this theory.

The third point depends also upon accepting the simulationist framework. Unless you have already accepted that because of the context one cannot ascribe the twin belief to John(te), the force of the point is lost. Since the twin belief could not have been caused by anything in his environment, the belief that ‘it’s raining’ is thus a very odd thing for John(te) to hold. If we ascribe beliefs through simulation and we ascribed this belief, it would not have been to explain the behaviour of
grabbing a bucket, but it would have explained some other behaviour like running for help. Thus, I claimed that such a belief would cause a different behaviour. If, however, one begins with ascribed beliefs on the basis of a theory, rather than from within a context, and posits something as being the belief, if that thing is physically the same in the twin case and the actual case, then, unlike the case with simulation theory, there is no alternative route to the claim that they could cause different behaviours. If we are talking about the same entity, we must be talking about the same causal powers.

Our final point about the non-individualistic nature of behaviour is consistent with the simulation view and its focus upon subjective experiences. If we do, in fact, distinguish between behaviour and beliefs, then regardless of whether there exists some in principle method for doing so, that is how we understand one another. Recall again the analogy with the case of induction. We’re still not sure how we distinguish some regularities from others and thus, we’re not sure when induction is justified. Nonetheless, we use induction and, for the most part, apply it in the right circumstances. With beliefs, we somehow perceive the behaviour as distinct from its cause. This is true even though we can legitimately describe behaviour in a fully cause-inclusive manner. How we are able to distinguish the cause from the behaviour itself is not clear, but it’s clear that we do. If I am to construct a theory I must know how this
is done. If I simulate you, I simply do it the same way. The lack of being able to identify behaviour in a non-context-bound manner is only a problem for the theory theory.

It is part of the theory theory that supervenience cannot be violated. This is a result of the fact that beliefs are taken to be entities that are posited to explain some objective phenomena. If supervenience is violated, realism fails. I have been arguing that supervenience is in fact violated on the simulation theory but that since, on this theory, the contents of beliefs are still causally efficacious, the violation of supervenience does not imply a rejection of realism. Beliefs cause behaviour. Change the contents of beliefs and you change the behaviour. This is true on simulation theory even though brain states may remain the same.

The problem with Burge was that he didn’t offer any alternative perspective from the theory theory of folk psychology from which to judge his ideas that externalism and realism are compatible. He didn’t explain how it could be possible that although the contents of beliefs are context dependent they are still legitimate explanations of behaviour. It seems to me that in order to claim that they are legitimate explanations of behaviour a demonstration of their causal efficacy was required. I attempted here to demonstrate this through examples in which substitutions of contents which were externally determined and
did not supervene on brain states were prohibited. My claim was that if the behaviour cannot be explained equally well by either content, then the contents do not have the same causal force. One explains the behaviour while the other does not.

Accepting this line of reasoning, however, depends upon accepting the alternative conceptualization of folk psychology wherein we do not theorize regarding another’s mental states, but we understand them by, in a sense, experiencing them ourselves. I argued in the last chapter that we should be open to the suggestion of a new perspective for several reasons. Firstly, if folk psychology is theory, the externalist arguments demonstrate that it’s a false theory. Second, even if it’s false it seems like it would be practically impossible to give up our folk psychological notions. Thirdly, the concept of belief is perhaps best treated as analogous to the concept of induction in terms of its resistance to both justification and abandonment and, this being the case, perhaps we should follow the example set by Goodman and accept that in describing what we do we are, in some sense, justifying our use of belief. Finally, I argued that since the notion of behaviour is non-supervenient in the same way as the notion of belief, if folk psychology is a theory it’s not clear what it is supposed to be a theory of.

In this chapter I hope to have shown that if simulation theory is correct, then the externalist considerations do not pose a
threat to realism in psychology since it is built right into simulation theory that our attributions are context bound. There remains, however, a need to clarify the issue of how we can maintain realism if we are granting that belief contents are not token identical with any physical states or entities.

A good illustration of how the acceptance of the simulation theory will allow us to maintain realism is to be found in Burge's lung and respiration example, which was discussed in chapter two. Let's consider the example from the perspective of what I shall call "Sister Earth". Sister Earth is the same as Twin Earth except that it exists 100 years in advance of Earth. The example will also require that we assume that we are living on Earth in a time in which people advanced hypotheses about the human anatomy, but had never cut open a body.

In his example, Burge points out that the sub-events of respiration are not individuated in such a way that they are supervenient on the chemically described events that compose them. The way we individuate the lungs depends partly upon the environment in which the physical events take place. So if our lungs, that is, the physical entities we call our lungs, were put into another kind of body and served a different purpose, say a digestive purpose, we would not call them lungs. The fact that we individuate lungs in this way, however, does not lead us to believe that the causal powers of the lungs are affected by
our individuation. This, I argued, is because we have another way of individuating the lungs. We can point to them and we can describe them in physical terms. We know they exist. We know that they have certain physical properties and that those physical properties are responsible for their causal properties. It doesn’t matter what you call the physical thing or where the physical thing is located, if the physical properties are the same, so too should the causal properties be the same.25

Let’s assume that lungs, like beliefs according to the theory theory view, are posits in a theory. The theory is a theory about breathing and claims that lungs are causally responsible for bringing air into the body. But we must assume also that we have never seen lungs and we can’t be sure that there exists some physical entity or entities that correspond to what we think of as ‘lungs’. Our only understanding of lungs is in terms of their causal role.

Now we go to Sister Earth and find that although everything is exactly the same physically as on Earth, Sister Earthling science has advanced further than ours. They have performed surgery and autopsies. We ask whether they discovered anything corresponding to our idea of lungs. If they say ‘yes’ and point

---

25 Of course, the causal properties may not have an opportunity to be exhibited in a different environment. The important point is that the fact that the same physical thing would be named differently in another context, does not affect its causal powers in this context.
out some physical thing, we will feel that our theory has been vindicated. We will have good reason to be realists about lungs. But now assume they say 'no' and tell us that there is no separate system responsible for taking in air but that it is simply absorbed through the skin, for example. They call the process 'epideration'. It seems to me that if this were the case we might be inclined to think either that there must be some physical differences between ourselves and Sister Earthlings or that our theory was mistaken and there are no such things as lungs and respiration. What we wouldn't think, it seems to me, is that 'respiration' and 'epideration' are simply two different, context-bound, ways of referring to the same thing.

According to the view that folk psychology is a theory, we are currently in the same state with regard to beliefs as we were in the above story with regard to lungs. We have posited them as causes. In order for our theory to be vindicated it must turn out that there is something physical that matches up with the beliefs we posit that can actually physically cause the things that we are assuming they cause. But according to the simulation view, we are in the same state with regard to beliefs as we currently are with regard to lungs. This must be the case in order for Burge's analogy to work. And in order for us to be in the same state with regard to beliefs as we currently are with regard to lungs, we must have some independent way of
identifying beliefs. The simulation theory is based on this assumption.

The simulation theory claims that our ascriptions of beliefs to others is based on simulation rather than theorizing. We understand what others are thinking by experiencing the thought. Burge argued that individualists have conflated causation with individuation. But, I have argued, if you accept the view that beliefs and desires are posits in a theory, then we have no means of identifying them aside from causation. If you change the belief content which has been posited as the cause of behaviour, you change the cause. Thus conflating individuation and causation in the case of beliefs, as they are viewed by the theory theory, is not a mistake. If, like the imagined lungs case, there is no other way of identifying the entities, then individuation by causal factors must be absolute, not relative to environment.

The simulation theory does not view beliefs and desires as simply posits in a theory. If the simulation theory is right, we do have a means, independent of descriptions of causal powers, of identifying beliefs and desires. We identify them through first person experience. If we do in fact have a means of identifying psychological states independent of their causal properties, the idea that our identification of those states via our descriptions of content can vary with context is no longer
problematic. We can identify the psychological state simply by having it.

For example, this state I’m experiencing now is what I will call ‘the belief that P’. My doppleganger, who is in exactly the same brain state and is experiencing exactly the same thing as me calls her state the ‘belief that Q’. Realism about psychological states requires that my belief that P exists and causes me to behave. Let’s say that my belief that P is the belief that it has begun to rain. This belief then causes me to get up and close the window. My doppleganger’s belief is that it has begun to twrain and this causes her to get up and close the window. If it’s false, however, that my doppleganger and I have the same belief even though our brain states and, hence, our causal properties are the same, how can the belief that P be an explanation of my behaviour? If the simulation theory is right, it can be an explanation that is relative to a context.

If we already know that an entity exists, we can individuate it by reference to its causal powers which can be described differently depending on the context. This is quite different from positing the existence of something in order to account for some other phenomenon. For example, if I am sitting in a room with a man named John who begins to sing, I can identify John as the man sitting across from me or the thing that is now causing music. If I am in the next room instead, and I faintly hear some
music, I could posit a man named John as being the cause of the music or I could posit a radio, or a woman with a low voice, etc. It seems to me that, if we accept the simulation theory, our position with respect to beliefs is like the first case where we can see John and we can also pick John out with a causal description. If we accept the theory theory, our position with respect to beliefs is like the second case where we aren't sure whether a man named John is in the next room or even whether there is a man in the next room, but a man in the next room seems to be the best explanation of the phenomenon we are trying to explain. We aren't sure whether beliefs exist and are the cause of behaviour, but that seems to be the best explanation.

Our theory about the cause of the music will be vindicated if, when we look in the next room, we find a man named John. Our theory of beliefs will be vindicated if, when we look in the brain, we find a physical equivalent to the belief that P. If we look in the next room and find a radio, or even a man named Sam, our theory will be proven false. We cannot simply say that what we meant by 'John' was a relativized description of whatever physical thing was the cause of the music. In the same way, when we don't find a physical equivalent to the belief that P, it is not legitimate for us to then say that by 'the belief that P' we were offering a relativized description of whatever physical brain state was responsible for the behaviour in question.
If the simulation theory is right, we don’t posit beliefs, we experience them. This is also the case where John is sitting across the table from me. I don’t posit John’s existence; I experience John’s existence. Then I experience John singing. With beliefs, I experience them myself and then, when explaining your behaviour, I experience them in an off-line manner through my simulation of you.

I can individuate the cause of the music in a variety of ways. I can say that it was the man sitting across from me, it was the man in the room with me, it was John, or if we are in Slovenia, I can say it was Jonko. I can also individuate the belief in a variety of ways. For example, the explanation of your behaviour can be individuated by ‘the belief that it’s raining’ here on Earth or ‘the belief that it’s twraining’ on Twin Earth. As long as the off-line experience is the same in both cases, we know we are talking about the same thing. We know that a variety of descriptions fit John, the man sitting across from me and we know a variety of descriptions fit the belief that P, the experience I’m having now. It is, however, only because we can see John and only because I experience the belief that P that we know to what these other descriptions refer. Thus, if we already know that beliefs exist, we can individuate those beliefs using descriptions that are relative to a social or historical or environmental context. Burge’s claims make perfect sense from this perspective. There is no problem with the violation of
supervenience of belief contents on brain states just as there is no problem of the violation of supervenience of the name 'John' on John, because we have evidence other than a name for the existence of John and we have evidence other than a description for the existence of the belief that P.

There is, however, a certain disanalogy between positing a man, even positing a man named John, and positing a belief that P. We know what 'a man' is. We know that men exist and we know what kind of thing we are positing when we posit one. We also know that there are a variety of ways of referring to men. We can posit a man in the next room or we can posit an adult person who is not female or a male human being over the age of 18, etc. We cannot say the same about positing a belief. First of all, we cannot simply posit a belief. We must posit a belief about something. This is why using an example of a man named John, rather than just a man, is more appropriate. But the disanalogy goes further. If the theory theory is right, we do not, in the case of beliefs, know in advance what one would look like. Other than 'the belief that its raining' we have no way to refer to the belief that its raining. We think that it may be a brain state, but since we don't know what brain state it is, we can't say what other descriptions might also refer to this brain state. But this disanalogy simply makes the case in favour of simulation theory stronger and further demonstrates why the theory theory does require supervenience for realism.
The claim of individualism is that beliefs can't cause behaviour unless they are physically represented. What I have argued is that our explanations of behaviour that invoke these contents cannot be substituted for explanations which invoke alternative contents that supervene upon the same brain state. The efficacy of the explanation is lost once the contents change. In addition, the efficacy of the contents is explained by the efficacy of the brain states to which, within the given context, the contents refer. Unless, however, we already accept that beliefs exist as experiential entities, we cannot claim that we can refer to them with descriptions that are relative to a context. On the simulation theory, beliefs exist first as experiences and then as explanations. On the theory theory they exist first as explanations and then as entities. Only on the second type of view will the failure of the description of the explanations to remain the same in different contexts result in the denial of the existence of the entities. And only on the first view, wherein the existence of the entities is previously granted, can we legitimately use a variety of descriptions, that will vary with context, to pick out the entity.

On the simulation theory we begin with the belief contents rather than with theory and with positing beliefs. In fact, we can't even see someone as behaving without already attributing belief contents. Since we haven't posited a belief, but rather have experienced it within a context, substitutions based upon
physical realizations abstracted from context are not possible. The belief is never identified with a physical state in the first place.

If we treat theories within psychology as we treat theories within chemistry or biology, realism is an option. In the same way that we need to grant that bodies exist for biology to get off the ground, the existence of beliefs must be granted before we can even begin to talk about a science of psychology. We need to grant it in order even to test the hypothesis that the belief that P and the desire that Q caused behaviour X. Psychological realism is ruled out by externalist arguments only if one accepts the idea that folk psychology is a theory of behaviour which posits beliefs and desires to explain behaviour. We have, in the last two chapters, explored the alternative to this idea. The alternative is to reconceive folk psychology as non-theoretical.

If the analysis I have presented is correct, Burge's position is no longer internally inconsistent. One can maintain realism and externalism at the same time. And although Dennett has been labelled an instrumentalist, it seems to me that his claim to realism is justified, or at least possible, on this picture. Michael Friedman explains it the issues I have been confronting in this way:
...the problem with contemporary psychology (and philosophy) is not that it tries to give theoretical reductions or explanations of human cognitive activities. There is no objection to theoretical reductions in the abstract....The problem, rather, is the direction of the reduction. Contemporary psychology tries to explain individual cognitive activity independently from social cognitive activity, and then tries to give a micro reduction of social cognitive activity—that is, the use of a public language—in terms of a prior theory of individual cognitive activity. The opposing suggestion is that we first look for a theory of social activity, and then try to give a macro reduction of individual cognitive activity—the activity of applying concepts making judgements, and so forth—in terms of our prior social theory....From this point of view, the primary problem with contemporary cognitive psychology—along with much else in linguistics, philosophy of language, and philosophy of mind—is a commitment to individualism. Perhaps it is time we turned away from this commitment. (1981, 15-16)

What I hope to have demonstrated in the last two chapters is that our commitment to individualism has been based upon our commitment to the view that folk psychology is a theory. We have been reluctant to abandon individualism because of the fear that it would result in a denial of realism. This fear is only justified, however, within the context of the view that the attributions of intentional states to explain behaviour is theoretical. Once we accept that psychology does indeed have a subject matter, that intentional states were not simply made up to explain behaviour, individualism is no longer necessary. And, it seems to me, once we abandon the need for individualism we will be free to pursue much more fruitful avenues to understanding one another. Our science of behaviour will then legitimately include the study of social and historical
influences involved the attributions of intentional states. We know that brain states cause bodily movement. Psychology is the study of much more than that fact. Unless we abandon the theory theory view and individualism, however, it seems to me that we will be destined to continue to attempt to vindicate every piece of psychology we do.

26 Do extroverts complete name recognition tasks more quickly than introverts? Are people with indoor jobs more likely to be homebodies? Does age make a difference to satiation time on colour recognition tasks? Does education make a difference? Accepting psychology as a theory at the level of these types of generalizations is different from taking it to be a theory that, for example, extroverts exist. Or that a phenomenon called 'satiation' exists. In order to test these statements, one must assume the existence of extroverts and satiation. And in order to test whether my ascription to you of the belief that academics are generally lousy drivers is correct (which I would likely do by asking you), I must assume that beliefs exist.
Once it was pointed out that external factors must play a role in determinations of ascriptions of meaning and intentional state content, traditional views regarding psychological realism were threatened. Despite the attempts to uphold realism in the face of externalist concerns, three of which were considered in the second chapter, a realism that begins with the assumption that the states we attribute to explain and predict behaviour are theoretical constructs cannot succeed. It cannot succeed because it depends on supervenience to guarantee the reality of the states. If there are no grounds other than the causal, theoretical role that beliefs are taken to play to believe in the existence of beliefs in the first place, then the failure of the posited causal entities to supervene results in the denial of the existence of those entities.

I have argued that we can, however, maintain realism if we reject the notion that we are using a theory when we ascribe intentional states. If there is some independent means of identifying intentional states, then the failure of the content of the propositional attitudes to supervene on brain states does not lead to a denial of realism about the attitudes. It leads, instead, to the kind of picture that Burge drew in which the ascribed propositional attitudes individuate the cause of
behaviour and that cause can be individuated differently in different environments. Since the acceptance of the view I am putting forward depends, at least in part, upon the acceptance of an alternative view of folk psychology, however, we should review this alternative in more detail. This chapter, then, will be dedicated to the clarification and defense of simulation theory.

In introducing the debate between the theory theory and the simulation theory, Martin Davies and Tony Stone point out the three most troubling objections to simulation theory. These will be reviewed along with a fourth that was raised by Stephen Stich and Shaun Nichols. Taking the views put forward by Alvin Goldman and Robert Gordon as representative, Stich and Nichols present detailed arguments against the claims that have been made by simulation theory. The aim of this review is both to clarify further simulation theory and to demonstrate that, in addition to the fact that it would allow us to maintain realism, and in spite of the objections raised by Stich and Nichols and the objections discussed by Davies and Stone, there are other defensible reasons for maintaining simulation theory.

\[27\] Davies and Stone's discussion is found in their "Introduction" to *Folk Psychology: The Theory of Mind Debate*, Blackwell, Oxford, 1995a, 1-44. All of the other articles discussed in this chapter are, unless otherwise noted, also found in either the above or Davies and Stone, (eds), *Mental Simulation: Evaluations and Applications*, Blackwell, Oxford, 1995b.
Stich and Nichols begin their paper, "Simulation or Tacit Theory" by claiming that the proposal of simulation theory offers "a bold challenge" to the received view. Speaking of those philosophers supporting the simulation theory they say:

If these philosophers are right, two enormously important consequences will follow. First, of course, the dominant explanatory strategy in cognitive science, the strategy that appeals to internally represented knowledge structures, will be shown to be mistaken in at least one crucial corner of our mental lives. And if it is mistaken there, then perhaps theorists exploring other cognitive capacities can no longer simply take the strategy for granted. (1995b, 124)

The second consequence is the one that we have been more concerned with. It is that the central premise in the argument for eliminative materialism is based upon the notion that folk psychology is a theory and the idea that if it is not a theory, then it cannot be a false theory. Thus, if simulation theory is right, despite the objections raised by the eliminativists and the issues raised by externalists, psychological realism would still be possible.

At this point in their paper Stich and Nichols (hereafter abbreviated to S&N) have not yet begun to present their arguments. However, in his reply, "In Defense of the Simulation Theory" (1995a, 191-206), Alvin Goldman already takes issue with the characterization of simulation theory as presented by Stich and Nichols:
The implication seems to be that opting for the simulation theory over the theory theory would be an abandonment of the prevailing paradigm of cognitive science. And that sounds like mutiny. In fact, choice of simulation theory would not be a radical departure from familiar paradigms in cognitive science. It is true that a great deal of cognitive science fits the knowledge-rich paradigm stressed by S&N. But...[t]here is also a substantial tradition that posits knowledge-poor procedures, i.e. processes or heuristics that are relatively simple or crude, and do not depend on quite so rich a set of rules or so complex a knowledge base. The simulation theory fits this tradition fairly comfortably. (1995a, 192-193)

Goldman cites examples of inferential heuristics proposed by Amos Tversky and Daniel Kahneman in defense of this claim. ²⁸

Goldman has addressed and perhaps attempted to tone down part of S&N’s claim that if simulation theory is correct two enormously important consequences follow. That is, since knowledge-poor procedures are routinely invoked in cognitive science, it wouldn’t be shocking, at least, to find that we use such procedures to attribute intentional states. It would, of course,

²⁸ Kahneman and Tversky believe that rather than representing rules or knowledge structures as complex as probability calculus, we use heuristics. Goldman (p.77) discusses ‘anchoring and adjustment’ and ‘representativeness’. Rather than going through complex calculations, people may sometimes begin their probability estimates with an ‘anchor’, which is determined contextually, and then adjust upward or downward. The ‘representativeness’ heuristic is used to determine probabilities by considering the degree to which an event or property is representative of the class to which it belongs. Thus although it is not possible for any conjunction to be more probable than either of its conjuncts, subjects routinely violate this rule given certain types of scenarios. For example, subjects rate the probability of a particular woman, Linda, about whom they have been given details of her past, being a feminist and a bank teller as higher that the probability of her being a bank teller. (Tversky and Kahneman 1983)
still be of great importance to know that this is what we do rather than applying a theory, as has been assumed. The second consequence pointed out by S&N is not mentioned by Goldman.

That the central premise of the eliminative materialist arguments would be rendered irrelevant is, it seems to me, just as S&N say, an enormously important consequence. It is this important consequence that we have been addressing in this thesis. And although it not discussed by either Goldman or S&N, it seems to me that this consequence actually works as an argument in favour of simulation theory. Even if one’s ultimate goal is not to defend realism, but rather to determine the manner in which we make intentional ascriptions, the fact that simulation theory may enable us rationally to maintain our folk psychological ascriptions whereas eliminativism would require us to abandon them, should in itself speak in favour of simulation theory. Thus, although it certainly was not their intention, it seems to me that S&N have, in their introduction to simulation theory, actually listed one of the reasons we should be motivated to consider it seriously. After all, an explanation of what we do that does not result in the conclusion that our behaviour is irrational and based upon false premises is, ceteris paribus, preferable to one that does. As I said at the outset, however, there are other grounds for supporting the simulation theory proposal.
S&N quote Gordon, from an unpublished paper, as offering the following argument in favour of simulation theory over theory theory: "To apply the alleged common-sense theory would demand anomalous precocity". Thus they refer to this argument as the 'anomalous precocity' argument. They expand upon the point by also quoting Goldman:

Children seem to display interpretive skills by the age of four, five or six. If interpretation is indeed guided by laws of folk psychology, the latter must be known (or believed) by this age. Are such children sophisticated enough to represent such principles? And how, exactly, would they acquire them? One possible mode of acquisition is cultural transmission (e.g. being taught them explicitly by their elders). This is clearly out of the question though, since only philosophers have even tried to articulate the laws, and most children have no exposure to philosophers. Another possible mode of acquisition is private construction. Each child constructs the generalizations for herself, perhaps taking clues from verbal explanations of behaviour which she hears. But if this construction is supposed to occur along the lines of familiar modes of scientific theory construction, some anomalous things must take place. For one thing, all children miraculously construct the same nomological principles. This is what the (folk-) theory theory ostensibly implies, since it imputes a single folk psychology to everyone. In normal cases of hypothesis construction, however, different scientists come up with different theories. (1995b, 80)

S&N agree that the fact that children are able to acquire and represent the laws of folk psychology by this early age is an impressive feat. But, they go on to argue, this is not the only example of an impressive feat of this kind. They cite knowledge structures that underlie a child’s linguistic ability as similarly impressive but nonetheless generally accepted to exist.
This point is partially conceded by Goldman in his response to S&N (1995a, 193-194). However, it is conceded only that the complexity issue does not present a knock-down argument against the theory theory. He claims that, in fact, since S&N seem to agree that folk psychology is not acquired 'at mother's knee', which has been widely assumed, S&N are actually making a concession.\textsuperscript{29} In moving away from this model, we are moving away from what has standardly been assumed within the philosophy of mind for some time. As S&N grant, we are also moving away from a picture of the theory of folk psychology as being learned in a way similar to how children might learn history or chemistry or astronomy where theory construction is involved, toward some sort of internal 'special purpose' component. As Goldman notes: "This could take the form of an innate knowledge structure or a special-purpose learning mechanism. But it could also take the form of a special type of heuristic: the off-line simulation heuristic" (1995a, 194). It seems then, that although S&N show that this argument does not refute the possibility of the theory theory being correct, they do not demonstrate that it poses any problem for the simulation theory. Furthermore, some concessions are made in terms of rejecting the standard theory construction model of acquisition of folk psychology.

\textsuperscript{29}Goldman quotes Paul Churchland from \textit{Matter and Consciousness} as an example of the standard view: "All of us learn (the folk psychological) framework at mother's knee, (as we learn our language), and in so doing we acquire the common sense conception of what conscious intelligence is...It embodies the accumulated wisdom of thousands of generations' attempts to understand how we humans work" (1988, 59).
In their "Introduction" to *Folk Psychology: The Theory of Mind Debate*, Davies and Stone discuss the issue of when it is appropriate to view something as a theory. This discussion, it seems to me, is relevant to the present point under discussion about standard views on theory construction. They point out, drawing on arguments made by Simon Blackburn, that our conception of knowledge of theory must be strong enough to avoid collapsing into a notion wherein we claim that anytime we are able to do something, we make tacit use of a theory. Blackburn puts the view to be avoided in this way: "If we are good at something...then we can be thought of as making tacit (very tacit) use of some set of principles that could, in principle, provide a description of a device, that is also good at it" (1995, 275). Davies and Stone continue:

If our conception of (tacit) knowledge of a theory is as weak as this, then we face the prospect of an early collapse between the theory theorist and the simulation theorist. For, an advocate of the view that our folk-psychological practice is underpinned by an ability would have little difficulty in agreeing that there can be a theoretical description of a practical ability—just so long as the ability in question has structure to be articulated, which our folk psychological ability surely has....So, the claim that folk-psychology is a theory must be stronger than the claim that our folk-psychological practice can be given a theoretical description. (1995a, 7)

One of the ways to make the claim that folk psychology is a theory more robust is to propose an analogy between folk psychology and scientific knowledge, our paradigm case of theories. This analogy is what is commonly thought of when we
talk about folk psychology as being a theory. But we have just discussed one of the serious disanalogies. As Goldman pointed out, it cannot be the case that we acquire folk psychological knowledge in the same way that we acquire scientific knowledge. Folk psychology does not have to be actively taught. Furthermore, it does not appear to be subject to empirical evaluation in the way that other scientific theories are. Familiar platitudes such as, 'persons tend to feel pain at points of recent bodily damage', for example, are more like definitions than hypotheses. So, if it were the case that S&N relied upon the professional science analogy to maintain the robustness of the view that folk psychology is a theory, there would be some problems.

As was discussed above, however, S&N don't rely on the scientific knowledge analogy anyhow. Their attempt to avoid trivializing the notion of theory in the way that Blackburn describes is instead to draw on an analogy with linguistics. But Davies and Stone also point out some disanalogies here. The complexity of the rules of grammar, for example, seem to rule out the possibility of learning. The rules of folk psychology are, in comparison, very simple and perhaps amenable to some alternative explanation:

The complexity of the speaker/hearer's task seems to rule out putative alternative explanations, such as explanations in terms of general learning rules. But, in what sense would a folk-psychological theory show comparable richness and sophistication? Here,
there seems to be a serious disanalogy with linguistics. For folk psychology is often thought to have a simple and homely content (hence, the epithet 'folk'). Once its principles are articulated, they seem to be obvious. (1995a, 9)

Notice that this is just the opposite point that Goldman made to which S&N introduced the analogy with linguistics in the first place. It looks as though folk psychology is complicated and sophisticated enough so that we would question whether a child of four or five could possess the knowledge required to make predictions about and offer explanations of behaviour, but it is not so complicated that an understanding of the concepts involved requires the positing of a tacit theory. The concepts, once explained, could, it seems, likely be readily understood by an older child or, at least, a young adult. This is not the case with the rules of grammar, which are much more complex. But does the level of complexity really matter? Isn't the issue just that if the principles of folk psychology are too complex for children to learn one could claim that they could be held tacitly in the way that the theory of grammar is?

One could respond that it could be the case that the theory of folk psychology is held tacitly in the same way as the theory of grammar, but, since there is a vast difference in the level of complexity of the two theories, our rationale for making such an assumption is weaker in the case of folk psychology. This weakness of rationale, which is a result of the differences in complexity, presents, it seems to me, an important disanalogy
with grammar. The disanalogy is important to Blackburn's point. If we argue that the claim that folk psychology is non-trivial is justified through the analogy with grammar, then if complexity is a prime motivator for positing a tacit theory in the case of grammar, we should expect this to be the case with folk psychology. But, at least prima facie, this doesn't seem to be the case.

Gordon makes the further, stronger point that the lack of complexity is very important for folk psychology since, according to the theory theorists, our competence in predicting behaviour is explained by folk psychology—the principles that underlie our mental attributions and explanations of behaviour:

Theory theorists generally assume that the laws of folk psychology can be formulated in terms of the common-sense mental vocabulary. They must assume this if they hold, as most seem to, that the common sense theory implicitly defines these terms. Generative grammarians, on the other hand, have no such compunctions about introducing a technical vocabulary....Analogy: It's all right to say that to catch a ball children have to solve differential equations, as long as this doesn't entail that they have to understand differential equations. (1995b, 177-178)

Thus the lack of complexity of the vocabulary of folk psychology is no accident. Grammarians have no problem in introducing a technical vocabulary since they do not claim that what explains our competence in grammar is a folk grammar. But theory theorists cannot introduce technical terms since doing so would conflict with the idea that our understanding of behaviour is
based upon an understanding of a theory, and many people, especially young children, could not understand a more technical theory. The question is whether young children of four or five years can really understand the theory at the level of technicality at which it now stands.

S&N could respond to Gordon’s line of argument simply by claiming that they are not representative of ‘most theory theorists’ and do not accept that an understanding of the concepts of folk psychology is required to predict and explain behaviour. They could claim that although the child of five doesn’t understand the concept of belief, because he has no access to the tacit theory by which the concept is defined, he is nonetheless appropriately utilizing the concept on the basis of his tacit theory. As Davies and Stone point out, however, "...given the distinction between having a body of knowledge and using it, a theory theory also needs to provide an account of the information processing mechanisms that enable that body of knowledge to be put to use" (1995a, 15).

Let’s summarize the lines of this argument so far. Gordon and Goldman make the argument that children would have to possess ‘anomalous precocity’ to apply the folk psychological theory. S&N argue that the requirement of applying a complicated theory is not as outrageous as it sounds especially given the example of grammar. The theory of grammar is highly complicated, yet we
accept that our grammatical behaviour is based upon a just such a tacit theory. Folk psychology, S&N argue, could also be a tacit theory. But while the complexity of grammar seems to force a view of a tacit theory, the case is not as compelling with folk psychology, given its relative lack of complexity. In addition, most theory theorists not only rely on folk psychology to explain our ability to predict behaviour, but view that ability as dependent upon an understanding of the concepts like belief. And since it is the theory that defines the terms of the mental vocabulary, like belief, an understanding of the theory is necessary.

S&N might respond to this by distinguishing between understanding folk psychological concepts and engaging in folk psychological practice, claiming that it is only the latter that is of concern and that the latter does not depend upon the former. Thus, one can use the concept of belief without understanding the concept of belief. The discussion concluded with the note that if this is the position of S&N, they still owe us an account of how the tacit knowledge is accessed. How can the child, who doesn't understand the concept of belief, be successful on a false belief task? In addition, since the analogy with grammar has its limitations, S&N have not demonstrated how the view that folk psychology is a theory offers a more robust sense of theory than the trivial sense of theory that Blackburn discusses.
It seems to me that in their response to Goldman and Gordon, S&N have not closed the door on the issue of anomalous precocity. In fact, as we have seen in our discussion, a number of other issues and questions for the theory theory are raised in response. The same thing can be said, it seems to me, of the next issue, which was also raised by S&N. Listed as an argument in defense of the theory theory, S&N claim that our predictions and explanations of behaviour are 'cognitively penetrable'. What this means, for S&N\textsuperscript{30}, is that they are sensitive to what we know and don't know about the laws that govern a system. S&N claim that:

If we are making predictions on the basis of a set of laws or principles, and if there are some unexpected aspects of the system's behaviour that are not captured by our principles, then our predictions about those aspects of the system's behaviour should be less accurate. Theory based predictions are sensitive to what we know and don't know about the laws that govern the system; they are cognitively penetrable....If we can find cases in which the ignorance about the workings of one's own psychology leads people to make mistakes in predicting what they, or other similarly situated people, would do, it will provide yet another reason to think that the off-line simulation theory is untenable. (1995b, 150)

\textsuperscript{30}The term was actually introduced by Zenon Pylyshyn. His meaning, it seems to me, is a little different. In his Computation and Cognition: Toward a Foundation for Cognitive Science, he says "The rationally explicable alterability of a component's behaviour in response to changes in goals and beliefs is what I refer to as cognitive penetrability" (1984, 133). Thus, if a component's behaviour is altered by new beliefs then it is penetrable. If the system is 'dumb', non-penetrable or, as Fodor (1983) says, 'encapsulated', then it proceeds to do whatever it does regardless of any new beliefs or goals that the larger system may have. The differences between Pylyshyn's meaning and S&N's don't, however, matter for our present purposes.
S&N then claim that such cases are common. They cite three examples using experiments conducted by Nisbett and Ross. In all three cases we are asked to imagine ourselves as subjects in studies that were actually carried out by Nisbett and Ross.

In the first study, under the pretense of taking part in a consumer opinion survey, subjects are asked to choose a preferred article of clothing from a table that has set out a number of the articles. Nisbett and Ross found that, although all the articles of clothing were actually identical, subjects tended to choose the right-most article as preferable (Nisbett and Ross 1980, 207).

In the second example we are to imagine that someone in the office is selling lottery tickets for $1.00. Some people who buy tickets are simply handed a ticket. Others actually choose the ticket they want. It turns out that, the next day when the ticket seller attempted to buy the tickets back, those who chose their own tickets demanded a much higher price for them than those who were handed their ticket. Those who had no choice in their ticket asked an average of $1.96 while the choice subjects asked an average of $8.67 (Nisbett and Ross 1980, 136).

The final example cited by S&N concerns the demonstrated fact that once a subject is convinced that she has a particular trait, she will continue to believe this even in the face of
overwhelming evidence to the contrary. Thus, once a subject is
convinced by an experimenter that she has the ability to tell
fake suicide notes from real ones, for example, she will
continue to believe this even after she is told that she was
simply the subject of a study and that she does not really
possess this skill. In addition, third-person observers also
exhibit this 'belief perserverence' and do so to an even greater

According to S&N, if the simulation theory were correct the
results of these studies would not be surprising to us, but they
are surprising to us. If we really did use simulation when
attributing intentional states, then when we imagined ourselves
in the situations Nisbett and Ross describe, that is, when we
simulate the circumstances described by Nisbett and Ross, we
should react in the same way as the subjects of the studies. We
should accurately predict what the subjects do. But, in fact, we
do not. S&N claim that we do not because we base our judgements
on what the subjects will do on our folk psychological theory.
If our theory is wrong, then our judgements will be wrong.

In his reply Gordon cites two objections to S&N's line of
argument. There are methodological problems, which Goldman
(1995a, 202-203) also discusses, and there is the problem of
what Gordon calls 'imaginative impenetrability' (1995b, 175-
To illustrate his point he uses the analogy of imagining the Muller-Lyer illusion:

...visualize two straight lines of equal length, one just above the other. One of the lines has a regular arrowhead at each end, the other has an inward pointing arrowhead at each end. Now, which line is longer?...There are two problems. One is methodological: It's silly to ask 'Which line is longer?' because I stipulated that the two lines are of equal length. Suppose we ask 'Which line looks longer?' There is still no reason to think we'll get the illusion going. That's because of the second problem which may be called the problem of imaginative impenetrability: Even if visual imagery is a product of the visual-perception system running off-line, it is surely going to bypass some stages of visual processing, especially early stages--and the illusion may originate at one of these stages. (1995b, 175-176)

Thus, the methodological problem with the suggestion of S&N that we imagine ourselves in the studies set up by Nisbett and Ross is that we are given either more or less information than the actual subjects in those studies. We are told, in the first example, that the clothing garments are identical. This is analogous to the Muller-Lyer case in which we are told that the lines are identical. It may be that when we are asked to imagine the situation, we bypass a particular stage of cognitive processing and it is this stage that accounts for the preference to chose items on the right.

In the second example, involving the lottery tickets, Gordon argues that the imagined situation is just too far from the real experimental condition and, thus, once again, the methodology is flawed. In the experimental situation one experiences either
choosing the ticket or receiving a previously chosen ticket, a
time lapse takes place and then a decision is made about the
price when selling the ticket. In imagining the scenario we do
not experience this time lapse nor do we experience choosing or
receiving a previously chosen ticket. In addition, S&N ask us to
imagine ourselves in both roles, the role of a person who chose
a ticket and the role of a person who did not. We are then asked
to compare them. Goldman points out further discrepancies. He
notes that if you did chose the ticket you would likely have
chosen the numbers you did for a particular reason:

In a real case there would be tickets with actual
numbers, and you would deliberately reflect on those
numbers and possibly come up with some 'grounds' for
preferring one of the numbers to the others. But if
you casually imagine yourself in this situation, you
would not exert the same deliberation and
reflectiveness. (1995a, 203)

In the final example we are told at the outset that the
information we receive about ourselves, such as being good at a
particular task like telling fake suicide notes from real ones,
is false. Since we don't ever believe the information one can
understand why it wouldn't persist. This example presents both
methodological problems and problems of imaginative
impenetrability. The methodological problem is that clearly
beliefs can't be expected to persevere if one is told
immediately that they are false and the imaginative
impenetrability problem is that the design assumes that the
mechanism that is responsible for perserverence of beliefs must also be triggered by merely feigned beliefs (Gordon 1995b, 177).

The truth of simulation theory does not require that we are correct in our assumptions about the outcome of certain experimental situations. This will be true of course in situations where there is some inherent methodological problem with the expectation that we should be able to simulate a situation, but it is also true because simulation, but not simulation theory, has its shortcomings. As is pointed out by Gordon, there are some types of psychological phenomena for which the difference between reality and simulation is crucial:

To imagine drinking six martinis doesn’t make one drunk, not even off-line. Here simulation is inadequate, and an independent source of knowledge is indispensable. That’s a shortcoming of simulation, but it isn’t a shortcoming of the simulation theory. It is a virtue of the theory that simulation fails to predict phenomena that our common-sense belief-desire psychology fails to predict. (1995b, 176)

S&N’s examples, then, do not refute the truth of simulation theory. In fact, they may even provide some support as is suggested above by Gordon. If the information provided to the subject is faulty, then the input to the decision making-processes (running off-line) will result in an incorrect answer. In order for the answer to be right, the inputs which are determined by information must also be right. In addition, the format in which the information is presented will also make a difference.
A further point on this issue is made by Jane Heal. There sometimes exist influences on decision making which cannot be captured by simulation. Heal says: "Replication (simulation) theory must allow somewhere for the idea of different personalities, for different styles of thinking and for non-rational influences on thinking" (1995, 48). As we have just seen, it is not the case that simulation theory requires that everything that affects decision making in a particular person can be simulated by another. And, just as we would expect, the cases where we cannot simulate, such when a person has taken a particular drug, will be examples of cases where we fail in our predictions of behaviour. In addition, simulation theory does not need to reject the notion that we can supplement our ability to simulate with some pieces of theory. This does not, however, undermine the simulation theory/theory theory distinction. This issue of individual differences will be discussed further below.

There are a couple of further points I would like to raise before leaving this issue of cognitive penetrability. Firstly, it seems to me that accepting cognitive penetrability has some consequences for S&N’s position on the last issue we discussed: anomalous precocity. According to S&N, what it means to say that the theory of folk psychology is ‘cognitively penetrable’ is that our predictions of other’s behaviour are sensitive to what we know and don’t know about the laws of psychology that govern the behaviour that we are attempting to predict.
Recall that the response of S&N to the claim by Goldman and Gordon, i.e., that predicting behaviour on the basis of a theory of folk psychology would demonstrate anomalous precocity, was to claim that since the theory is tacit, just like the theory of grammar, the fact that it is complex does not preclude four year olds from being able to use it. The theory, we have assumed, is made up of generalizations such as, 'If someone desires X and believes that Z will result in X, then, everything else being equal, that person will bring it about that Z'. According to the Nesbitt and Ross examples, according to S&N, there are some generalizations of folk psychology that are simply not represented in our tacit theory. Our failure to predict certain outcomes is based on a failure of our tacit theory. But why would a generalization like the one stated above be represented and a generalization such as 'given a choice between several identical objects, people will choose that object which is right-most' not be represented?

S&N still owe us an account of theory acquisition. They have rejected the scientific model of hypothesis testing and have opted for the analogy with grammar. A scientific model would account for our ability to answer questions correctly regarding some psychological principles and not others. On the professional science model, we may simply not have encountered any situations where we would test an hypotheses regarding how people will choose given an array of identical objects. We
acquire our knowledge through testing and confirming hypotheses. But on the grammar model, we still require an explanation of why we use or access some generalizations and not others, or of why we would represent false generalizations.

If S&N's view is that there is some kind of cognitive module\textsuperscript{31} at work such as an innate knowledge structure or a special-purpose learning mechanism,\textsuperscript{32} then we need to know just what is innate and why or how such a special-purpose learning mechanism would work to represent some pieces of folk psychology and not others. Our failure to simulate the subjects in the Nesbitt and Ross experiments is, it seems to me, easily explained. On the theory theory account, however, it is not so easy to explain why we have tacitly represented some psychological generalizations and not others.

The focus of my second point remains upon the issue of the analogy of the theory of folk psychology with the theory of grammar and our failure to predict the results of the Nesbitt and Ross experiments. S&N claim that "[w]hen our folk psychology is wrong, it is to be expected that our predictions will be wrong too" (1995b, 152). The fact that our folk psychology can

\textsuperscript{31} This would be odd given that they are arguing that folk psychology is cognitively penetrable. This would be where it would be important to distinguish their meaning from Pylyshyn's.

\textsuperscript{32} These options, which I cited earlier in this chapter, are suggested by Goldman, along with the option of the off-line simulation heuristic (1995a, 194).
be wrong, it seems to me, presents another quite serious disanalogy with the theory of grammar. Although we do indeed sometimes make grammatical mistakes, this is not because our theory is wrong.

In studying the rules of grammar, linguists study what we do and then attempt to determine what the rules might be behind what we do. Of course, some of us sometimes make mistakes, but if there is something that we all do consistently with language it cannot be considered wrong. Instead it must be explained by the rules. If we all thought that the string of words ‘Yellow yesterday went did outside’ constituted a grammatical sentence, the conclusion would not be that we all hold a false theory of grammar, but that the theory of grammar that we hypothesize that we tacitly hold must explain the fact that we find the string of words grammatical. Furthermore, although individuals make mistakes on occasion, these mistakes are explained in terms of some kind of cognitive or performance malfunction brought on by fatigue or illness or brain damage, etc. Since the theory is developed to explain what we typically do, it is not the theory that is responsible when individual errors are made.

The exception is the case of young children who make grammatical mistakes. It appears, for example, that all children make similar kinds of errors such as overregularization. When the child first begins to use the past tense, he or she will begin
adding 'ed' to irregular verbs and will start saying things like 'he goed to the store' or 'she doed it wrong'. And, in these cases, it seems that children are using a theory. In fact, the rule of the theory seems to override explicit instructions from parents to disobey the rule. So when parents attempt to teach a child the correct grammar by asking them to imitate what they say, the children tend to still use the rule. The problem still, however, is not with the theory, but with the application of the rules of the theory. It is, again, a performance problem. As they grow older and are exposed to more examples of proper application of the rule, this grammatical error disappears. And, if it didn't, if we all continued to say things like 'she goed to the supermarket', that sentence form would be the one our theory would have to explain and it would not be considered a mistake.

This does not appear to be the case with the theory of folk psychology. It seems that, in the case of folk psychology, our tacit theory can contain principles that are simply wrong but that, nevertheless, everyone holds. If our theory of folk psychology is tacit and we all somehow come to represent the same theory, it is unclear how or why we all come to represent the same false information. It seems to me that the view that folk psychology is a theory loses some of its force if we claim both that we need to posit a tacit theory to explain our competence and we need to posit a tacit theory to explain our
incompetence. Or, at least, the issue of acquisition of the theory becomes even more pressing and the analogy with grammar is difficult to sustain.

Another common objection to the simulation theory that is discussed by Davies and Stone (1995a, 18) is that in order to drive the simulation one must have knowledge of the object he or she is simulating and that knowledge must be organized in something like a theory (Daniel Dennett 1987, 100). Dennett compares imagining himself to be a bridge and wondering how he would react to the wind blowing with mental simulation, which he describes in terms of 'putting oneself into somebody else's shoes'. Dennett argues that in order to determine how the bridge would react, he needs some theory pertaining to the engineering of bridges. Davies and Stone consider this issue in terms of a distinction between theory-driven simulation and process-driven simulation.

Simulation, it is conceded by Goldman (1995b, 85) can be theory driven. An example given by Davies and Stone is of a computer simulating the economy of a country. The computer does not go into a recession or anything off-line isomorphic with a recession. But mental simulation does not work in this theory-driven way. It is driven by process rather than theory. In the case of mental simulation, if the inputs are the same or relevantly similar, then, without any knowledge of theory, the
simulator can come up with the same, or relevantly similar, end states as the system because he or she goes through a similar process. Thus, although it's true that in order to know how a particular bridge would react to a certain wind force I would require some theory about physics and engineering of bridges, in order to know what you would do given that it's 2:00 a.m., you're tired, and you've just finished your novel, I don't have to know a theory of human psychology. I know that you will be going to bed. I simply input that information and run it through my own system. Of course, if I don't input all the relevant information, such as that you still need to prepare a lecture you will be presenting tomorrow, I will not come up with the right prediction of your behaviour. But, in this case, it isn't any theory of psychology that I'm missing—it's information.

The final objection we will consider is along lines similar to the last one insofar as it is a complaint about theory entering into simulation. Critics argue that in order to set the simulation up in the first place, one must take into account the differences between oneself, being the simulator, and the person whom one is simulating. Since there are always differences between oneself and the other, in order to decide which differences are relevant and which pretend beliefs and desires I should input, I need to use some theory. In order to know whether I should input a particular belief, I already need to
know whether it will be important to the outcome whether I input that belief.

The response to this, which Davies and Stone call 'the direction of gaze response' (1995a, 20), is that simulation does not look inward to discover the person's motives and states, but rather looks outward to his environment, from his perspective. When we try to understand a piece of behaviour we do not typically do so by trying to understand the person behaving. Rather, we try to understand his situation. Putting ourselves into his situation may not always work because there may be some things about his situation, about the world from his 'gaze' that we do not capture in our simulation. But accounting for these differences is different than accounting for some kind of 'inner' or 'personality' differences. If we fail to input relevant information, or when we succeed in inputting the relevant information, it is because we have either inadequately or adequately placed ourselves in his situation. Once that is done, the issue of relevance is solved by the process. Theory is not required. In fact this issue poses more of a problem for the theory theory view since it is not clear how theory could ever tell us what is relevant. This problem, in fact, has proven to be one of the most difficult for computer scientists in their
attempts to design a computer to model human decision making. The problem is known as 'the frame problem.'

Simulationists admit, however, that inductively based generalizations are sometimes used, but that they play only a secondary and heuristic role in predicting behavior. That is, we will sometimes use information that we might understand as 'personality traits', such as 'Max is a night owl', to predict behaviour like what time Max might go to bed. According to Davies and Stone, according to the simulationists: "Use of the inductively based generalization is just a shortcut--and a shortcut whose use is intelligible to me only derivatively upon engagement in mental simulation" (1995a, 21). Goldman puts it this way:

I am not saying, it should be emphasized, that simulation is the only method used for interpersonal mental ascriptions, or for the prediction of behaviour. Clearly, there are regularities about behaviour and individual differences that can be learned purely inductively. If Jones always greets people with a smile whereas Brown greets them with a grunt, their acquaintances can form appropriate expectations without using simulation....The suggestion, then, is that simulation is an intensively used heuristic, and one on which interpretation fundamentally rests. Inductive or nomological information is not wholly absent, but is much sparser

---

Jane Heal discusses the frame problem for the theory theory in "Simulation, Theory and Content" (1996, 81-85) She argues: "The theory-theorist is committed to the claim that we have--tacitly at least--solved an extremely important precursor to the Frame Problem in Artificial Intelligence, namely the problem of providing a general theory of relevance. And this claim is highly improbable" (p.81).
than the folk-theory approach alleges. (Goldman, 1995b, 83)

Thus, although theory does sometimes enter the picture, it is not necessary, as the critics argued, to get simulation going in the first place. Furthermore, it is not clear in what sense these types of inductively based generalizations would be constitutive of psychological theory or, at least, what would distinguish these types of inductive generalizations from any others. This type of theory does not involve any rules or principles of psychology. All that is required is inductive reasoning. Allowing for this type of theory does not pose any threat to the explanatory force of the simulation theory.

The simulation theory has strong explanatory force and, it seems to me, has posed some serious challenges to the theory theory view of folk psychology. I have here discussed the objections to the simulation theory that I believe are the most serious. There are, of course, many other objections to simulation theory. There are, for example, detailed exchanges over what the empirical studies demonstrate. It seems to me that, in at least one area of empirical study, the results once again favour the simulation theory. In fact, I will argue in the next chapter that the data from the studies that have been conducted with

---

34 The debates over the empirical results can be found in the same articles that have been under discussion here. See esp. S&N and Gordon and Goldman.
children with autism demonstrate the plausibility of the simulation theory over the theory theory.

For the purposes of this thesis, a defense of the simulation theory is required insofar as we are concerned to refute the view that beliefs and desires are posited entities in a theory. The aim is to uphold realism via the denial of theoriticity. Since, on the simulation view, our attributions of beliefs and desires are based upon a process rather than a theory, their existence is not open to challenges based upon theoretical considerations. Of course, as we have discussed, their veracity is still open to dispute on a case by case basis, and this is based upon whether the correct inputs were run through the process. But since the folk psychological states of belief and desire are what result from the process of simulation, the existence of these folk psychological states is never in question.

The theory theory view is that we understand one another on the basis of a tacit folk psychological theory. We have here considered two of the representatives of that view, Stich and Nichols. Their view of the theory theory is that it is more like the theory of grammar than a professional scientific theory. We found, however, that there are problems with this analogy and that there are general questions that remain for the theory theorists to answer, such as why the tacit theory includes false
principles, how we acquire the theory, and how we access the theoretical knowledge. As I now hope to demonstrate, in the next and final chapter, the problems for the theory theory do not end here.
Chapter Six

Autism and the Simulation Theory

This thesis has been an attempt to show that psychological realism can be maintained if the simulation theory is adopted, that other proposed means of defending realism don't work and that, aside from its making realism compatible with externalism, we have good independent reasons to believe that the simulation theory is right. In the last chapter we considered and found problems with the main objections to the simulation theory. At the same time, we considered what some of the problems might be, aside from its incompatibility with psychological realism, with the theory theory view. This final chapter will focus on what I take to be one more of the good reasons to favour the simulation theory. I will examine the evidence for simulation theory from the field of the study of developmental psychology, with a focus on autism.

Autism is a tragic disability that is diagnosed on the basis of a number of characteristics rather than on the basis of its etiology. There is not complete agreement in the field as to what the central defining characteristics are. Typically, however, definitions would include the following: impairments in social skills, communication and creativity and imagination abilities; an aversion to change; perseverative behaviour
including perseverative language and echolalia; sensory and perceptual abnormalities of various kinds (Wing and Gould, 1979; American Psychiatric Association, 1987; World Health Organization, 1987). Since studies have shown that children with autism also lack an understanding of mental concepts and are unable to attribute intentional states, there has been excitement over the idea that this particular deficiency may explain many or all of the others.

When those studying autism talk about a lack of a 'theory of mind' in people with autism they do not always intend to be committing themselves to the theory theory as opposed to the simulation theory of intentional state attribution. In many, if not most, cases the 'theory of mind' terminology is used as a synonym for the 'concept of mind' or 'folk psychology' or the ability to ascribe intentional states. Thus it is not the case that the issue has been decided in advance in this field. The use of the 'theory of mind' term indicates, rather, that the issue of whether what is lacking is best understood as a theory or as an ability or skill is not an issue that has been viewed as of central relevance, at least until recently. What has been viewed as relevant is the idea that perhaps it is this lack of

35'Echolalia' is simply the production of language which 'echoes' what the speaker has heard. For example, if you ask Johnny, 'How are you?', he would respond with 'How are you?'. Echolalia can also be more complex than this as when delayed echolalia is used. The environmental cues may trigger appropriate language because the child is echoing something that he has heard in the past in the presence of the same environmental stimuli.
understanding or ability to ascribe intentional states that underlies and explains much of the social, communication and behavioural abnormalities found in people with autism. But, even on this issue, there is not complete agreement. Some psychologists, for example, think that the failure of children with autism on false belief tasks is attributable not to a lack of ability but to a lack of motivation. They argue that these children do possess the ability or the theory to succeed at these tasks, but they see no reason for doing so. Others argue that while the lack of a theory of mind, or ability to ascribe mental states, is an important issue it is indicative of other central processing problems and is not itself the cause of the abnormal behaviour but is instead one of the results. However, even those psychologists who do not grant the lack of a theory/ability of mind a central role in explaining the abnormalities of autism tend to agree that it is highly significant that people with autism fail at false belief tasks and, for whatever reason, do not ascribe beliefs and desires to others to explain their behaviour.

Given this state of things in the field of the study of autism I am going to proceed as if it has been granted that the lack of a theory of mind/ability is at least partially responsible for some of the abnormalities of autism. Granting, then, that people with autism lack the capacity to ascribe intentional states and to predict and explain behaviour on that basis, the question I
will be pursuing is whether what they lack is best understood as a theory or an ability.

There are several reasons why I believe that people with autism lack an ability to simulate rather than a theory. The first is that I believe that in typical cases what we are using is an ability and not a theory. Since my arguments for this view have already been given in the last and other chapters I will not pursue this reason further here. Second, since simulation involves a type of pretending, the failure of people with autism to engage in pretend-play supports the view that what they are missing is an ability to simulate. This claim is not, however, as straightforward as it sounds since children with autism do not completely lack the ability to pretend. The ability is certainly deficient, but it is not completely absent. Since there are some who claim that the data from pretend-play does not support the simulation view, I will discuss the arguments and studies cited by Paul Harris to support this point and will examine an alternative proposed by Alan Leslie. Leslie's view, as representative of the theory theory, will be examined further insofar as it explains other deficiencies and tendencies in autism. The theory theory will be considered in comparison with the simulation theory with regard to their explanations of: variability on tests of mentalistic understanding; lack of deficiency in theorizing in non-mentalistic areas; autistic
preferences for narrowly focused fact-based topics; and autistic aversion to change in environment and disruption of routine.

Studiest examining pretend-play abilities in children with autism have shown that autistic children are less likely to engage in pretend play with toys compared with retarded or normal children of the same verbal ability (Sigman and Ungerer, 1984; Baron-Cohen, 1987; Lewis and Boucher, 1988). While, on its own, this may suggest a simple inability to imagine, other findings from the same studies suggest that this is not the case. For example, it was consistently found that prompting elicits more pretend play from autistic children. This could be taken as evidence for the idea that motivation, rather than ability to imagine, is lacking in children with autism. Perhaps the children are simply unmotivated, because of their lack of interest in toys and other people, to engage in pretend-play. But, as Paul Harris has pointed out (1993, 231), imagination is also lacking in areas that are clearly interesting to autistic children.

While a person with autism may show an interest in numbers or history, this interest is likely to be manifested in unimaginative and ritualistic ways such as studying calendars or memorizing dates. It is not uncommon for a person with autism who has an interest in dates, for example, to be able to tell you what day, in any year, a particular date falls on. For example, you could ask such a person what day Christmas fell on
in 1934 and typically within less than a minute, he could tell you. A child with an apparent interest in history might be able to tell you when certain historical battles were fought, but he likely would not be able to tell you about why they were fought. It is also common for autistic children to obsess about particular details of subject areas in which they show an interest. A child interested in cars, for example, may want to talk about nothing else but red corvettes and may repeat the same facts about those cars every time you converse with him. Being motivated by a subject area is thus not enough to enable autistic children to engage in the related activities in a creative, imaginative or playful way: "Thus although the motivational hypothesis implies that the apparent imaginative poverty of autistic children is confined to a well-demarcated content area, namely the area of human interaction, the evidence suggests that their difficulties are more pervasive" (Harris 1993, 231). Furthermore, as we shall now discuss, the extent of the pretend play that is elicited by prompting suggests that while the ability to imagine is not completely absent, it is still deficient.

It seems that when prompted, autistic children are able to pretend-play when the pretense involves object-substitution. For example, they have no problems treating a blue serviette as a swimming pool, or cotton wool as snow (Harris 1993, 232). They were also very accurate in assessing what dolls could see,
although they were not good at understanding false belief. These findings suggest that children with autism have difficulties with some kinds of pretend play but not others: "They can treat one object as another object, and they can treat a doll as if she were a person who talks, acts and even sees. On the other hand, they have difficulty in imagining a belief they do not share, whether that belief is to be attributed to a doll or another person" (Harris 1993, 232). It seems, then, that as long as the pretense doesn't involve intentional states, children with autism, given prompting, are capable of it.

Again, this finding may indicate that the problem is not that children with autism are unable to imagine or simulate but that since their inability is restricted to the area of intentional states perhaps what they are lacking is theoretical knowledge in this one area. Perhaps they are unable to pretend about intentional states because they don't have a theory of intentional states and they can't pretend about something of which they have no understanding. There is, however, an alternative interpretation.

The alternative suggested by Harris, and with which I agree, is that autistic children are deficient in their planning abilities. It is this deficiency that accounts for their lack of pretend play without prompting, their lack of ability to engage in higher order, including intentional state, pretend-play, and
their rigidity when they do engage in pretend-play. In explaining this planning deficiency Harris makes a distinction between two types of actions:

...actions that can be guided by standard or habitual schemas evoked by the current context, and those that must be guided by a specific plan specially formulated for the task...., some actions require that a familiar or habitual schema, normally evoked by the current context, should be set aside so that guidance can be achieved instead by a temporary, internally-formulated plan. Autistic children, it will be claimed, have special difficulties in over-riding external or habitual control in this autonomous, planful fashion. (1993, 233)

When a typical child engages in pretend play there is a shift from external to internal locus of control. Initially it is the cues in the environment that prompt the child to pretend. A doll and a stroller prompt the little girl to pretend that she has a baby, for example. But eventually, the internal planning takes over and the child sets up scenarios that are unrelated to her actual environment. She may, for example, pretend that she is in a grocery store with her baby and is purchasing food items, although nothing in the home environment suggests this scenario. She will use unrelated items such as a rubber duck to represent a box of cereal or blocks to represent a loaf of bread and a jug of milk. This shift from the environment controlling the planning or direction of the scenario to the child controlling the planning of the scenario is what is missing, Harris claims, in children with autism: "...the autistic child is more reliant on the schemas evoked by the current context, and has great
difficulty in deliberately guiding his or her behaviour or responses according to an internally conceived plan that over­rides those schemas" (p.234).

Support for this view is found in work that has been done with autistic children in the area of problem-solving. Studies conducted by Prior and Hofman (1990) and Ozonoff, and Pennington and Rogers (1991) showed that children with autism perform poorly on problem-solving tasks as compared with learning disabled children. The differences were accounted for by perseverative errors. Autistic children would tend to repeat a strategy that was successful on the previous task. If they were successful at sorting things according to colour they would attempt to do so on the next task even though the point of the task was to sort according to shape, for example. Autistic children performed especially poorly on the 'Tower of Hanoi' problem which requires thinking ahead to what will happen after one makes the next move. It involves, in other words, imagining a hypothetical situation and the consequences of one's actions in that situation. These failures demonstrate that autistic children have difficulty in shifting strategies from those that were initially suggested by the environment and that they are

---

36 The Tower of Hanoi problem requires the child to move disks from two other pegs onto a third in order of size to form a tower. In moving the disks the child must respect the rule that the larger disks cannot be placed upon the smaller ones if the goal state is to be achieved. The object is to achieve the goal state in as few moves as possible and, hence, success requires the ability to think ahead or plan.
unable to guide their current behaviour in terms of future imagined contexts.

In addition to these problem-solving tasks, support for Harris' view can be found in the wide-ranging epidemiological survey done by Wing and Gould (Wing 1978; Wing and Gould 1979). These studies show that despite some competence for make-believe, especially in structured settings, the play of children with autism remains inflexible and repetitive. This is true even when we take into account cases where the pretend-play appears to be quite sophisticated. Harris offers an example of an 8 year old girl who pretends to be a waitress but each time she pretends the scenario is exactly the same. The questions she asks and the orders she takes are exactly the same and the make-believe breaks down as soon as she leaves the table unless she is further prompted.

It seems then, that on a simple level and to a limited degree, children with autism are able to pretend. But once the pretense requires further sophistication in terms of flexibility and non-repetitiveness, or the shift of control of the scenario to internal, or the imagining of non-existent contexts, children with autism fail. These findings suggest the view that the child with autism is lacking an ability to simulate, which is required for planning, rather than a theory of mind.
Harris describes two different levels of simulation.

At the simpler level, the child holds the current, actual situation constant, but imagines a different intentional stance toward some aspect of that current reality. At the more complex level, the child does not hold the current context constant, but instead imagines a hypothetical situation that conflicts with what is currently known to be the case, and then proceeds to imagine a different intentional stance toward that hypothetical reality. (1993, 239)

The data shows that autistic children usually pass tasks based on the first level and fail those based on the second. Harris' distinction thus nicely accounts for the data on pretend-play and problem-solving tasks: "Scrutiny of the tasks that autistic children pass shows that they do not require the child to set aside the current situation and to imagine instead a conflicting hypothetical situation....By contrast, that conflict is apparent in the tasks that autistic children fail" (1993, 239). Thus, the second level of simulation is not required in any of the tasks that the autistic children pass but is required in all the tasks that they fail.

An alternative is found in what is perhaps the most prevalent view on what is lacking in children with autism that would explain their failures on false belief tasks. This is the 'metarepresentation' view developed by Alan Leslie (1987). While it seems that the simulation theory view is the most natural one when we attempt to explain both the imaginative deficiencies of children with autism and their failure on certain mental state
concept related tasks, Leslie’s theory theory view also ties together these two deficiencies. Leslie claimed that the normal capacity for pretence is a reflection of the normal ability to employ and develop a theory of mind. This led to the hypothesis that the mechanisms constituting this capacity were impaired in autism. When he talks about ‘metarepresentation’, Leslie is talking about an information-processing data structure which is not something that is typically consciously known to the subject. A metarepresentation consists of two components, the second one compound. The first component consists of who the agent is. The second is a specification of the relationship between the agent and a) an aspect of reality, and b) an imaginary situation. For example, if Johnny pretends that the glob of mud in front of him is a pie, Johnny is related to the proposition ‘it is pie’ by way of holding an attitude (pretending) that the proposition is true of an aspect of reality (the mud). Leslie’s theory about autism is that autism is an inability to produce metarepresentations. So people with autism cannot form mental representations of the sort described above. They cannot, in other words, have thoughts about someone having thoughts.

Leslie rejects the simulation theory. He proposes instead a modular theory involving an innate specialized theory of mind module. The module is responsible for the deployment of metarepresentations and contains the theory of mind information.
Setting aside the innate, modular, knowledge aspect of Leslie's view for a moment, it seems to me that his 'metarepresentations' are, in fact, not inconsistent with the simulation theory.

The inability to simulate is the inability to imagine something to be the case that is not in fact the case and to judge what the consequences of that imagined state might be. So in order for me to predict what Susan might do now that she has realized that the pot on the stove has boiled over, I must be able to imagine what I would do, imagining that I am Susan, if the pot on the stove were to boil over. In doing this I can predict that she will run to the kitchen and take the pot off the heat. Of course, I need not do any of this consciously. But in order for me to do it at all, I must represent Susan and the pot boiling over and myself being Susan and the pot boiling over. Whether this constitutes metarepresentation is not clear. It seems to that, at least in Leslie's sense, it does not. But for now, let's grant that it does constitute metarepresentation of a sort. I must, then, be capable of metarepresentation.

I don't think that Harris' arguments, which focus upon a lack of planning abilities and the inflexibility of children with autism, are in conflict with this metarepresentation view. In fact, the second level of simulation, discussed by Harris, the absence of which he claims accounts for the failure of autistic children in certain tasks, might be viewed as another
Leslie himself claims that the metarepresentational process he talks about can be characterized as simulation, but that doing so adds nothing to the account. He believes that there is a specialized mechanism, a theory of mind module (ToMM), that deploys the metarepresentation early in development in normal children and that it is the failure of normal growth of this mechanism that results in children being autistic (Leslie and German 1995, 129). Why does Leslie think that the
metarepresentational system is better understood as theory than simulation? He claims that the ToMM hypothesis can certainly accommodate simulation, but that simulation is not entirely knowledge-free. It is because he believes that, in order to deploy the metarepresentations and ascribe intentional states, knowledge is required, that he believes in the existence of a tacit, innate theory of mind. Leslie claims that systems of representation themselves constitute bodies of knowledge: "To fully deploy such systems, additional abilities are required (e.g., inferencing that is sensitive to the structure of the representations)" (Leslie and German 1995, 128).

In characterizing Alvin Goldman's view, a simulation theory view, Leslie and German claim that Goldman's descriptions of simulation requires inferences to operate over metarepresentations. This, they claim, "makes it a less-than-radical, knowledge-and-ability account, where one of the abilities happens to be 'simulation'" (1995, 132). Leslie and German appear simply to assume that if one's account utilizes or accepts the possibility of metarepresentations and inferences that it has then abandoned the point of the simulation theory as an alternative. There are, however, two aspects of the theory theory vs the simulation theory to consider.

First, if we are using a theory then it's true that we will be making inferences based on information. But it's certainly not
clear that if inference is used to come to a conclusion that one is necessarily using a theory. In other words, while inferential reasoning is a necessary condition for something to be a theory it is not a sufficient condition. The other aspect of theories, the relevant aspect as I see it, is the fact that if one is using a theory to reach an hypothesis then one is accessing a store of theoretical information. For example, in order to predict that John will open the window, the theory theory says that we access the theoretical information that we have stored about what people will do given certain conditions, beliefs and desires, i.e., that the room is very hot and John appears to be uncomfortable and the window will open. Since John is uncomfortable and folk psychology tells us that, in general, people do not like to be uncomfortable and they try to alter circumstances in order to relieve their discomfort, we would conclude that John will try to relieve his discomfort.

According to the simulation theory, we do not need to access any such store of information. We simply simulate ourselves in John’s position and conclude that he will likely open the window. This conclusion is reached on the basis of the fact that this is what we would do ourselves. Since we don’t use any theory in our own case, it is unnecessary to use it in the case of others. Does this process involve inference? Well, perhaps insofar as the process involves inference in my own case, it does. If I am too warm, I may look around the room, see the
window and, in order to cool off, decide to open it. Have I used inference in this process? I have inferred that opening the window will make the room cooler, but I don't need any theory of mind to make this type of inference.

Admitting to the use of inference, it seems to me, does not make the simulation theory collapse into the theory theory. It does not justify Leslie and German's claim that knowledge is required in order to ascribe mental states. Sometimes, however, we will base our inferences on theory of mind or folk psychology. As Goldman points out, there are cases in which the theory theory is correct. For the purpose of this thesis, the main concern is whether the realism of the intentional states is established by a theory. If it is not, then whether we sometimes use theory in order to make intentional state ascriptions, is not a matter of concern. Furthermore, the fact that there may be cases in which we do use theory does not imply that the theoretical information is tacit or innate or modular. So in explaining the deficits of children with autism, Leslie's theory that the module didn't develop properly does not fare any better against the simulation theory even when the simulation theorists admit that there are cases in which theory may be used. If we don't need an innate knowledge store, why posit one?

While some simulation theorists would not object to the idea of talk of metarepresentations (Harris and Kavanaugh, 1993), there
are others who would. Furthermore those who don't object to talk of metarepresentaions still would object to the particular view that Leslie offers. A central aspect of Leslie's theory is something he calls the 'decoupling' of expressions. What he means by a 'metarepresentation' is a decoupled expression. Decoupled expressions disregard the semantic properties that those expressions would have in a normal representational context. This is why when we pretend something to be the case, we do not have the same expectations or make the same inferences as we would if it really were the case. For example, one may pretend that a banana is a telephone but not expect that he or she will hear any actual voices coming from it. The reason that pretending does not undermine the representational system and 'bring it crashing down'' (1987, 412) is because pretense involves metarepresentation where the inner symbols are decoupled from their ordinary meanings. Gregory Currie argues (1995), however, that the semantics of pretense do not require, nor should they receive, special treatment in order to avoid representational collapse. Furthermore, although Leslie's theory was, at least in part, developed in response to the fact that people with autism have difficulties with pretense and with understanding mental concepts, many people with autism can, in fact, metarepresent on Leslie's theory of metarepresentation.

Currie points out that, although we may not make the same inferences when we are pretending something to be the case as
when something actually is the case, the normal semantics of symbols are, at least in part, explanatory of pretend behaviour:

It is the fact that I imagine, exactly, that the pile of snow is a man, rather than a house, that explains why I stick a pipe in his mouth rather than attempt, say, to form a proto-roof for him. 'Man' means the same thing to me when I imagine the snow pile to be a man as it does when I think to myself, concerning a real man, 'That man is drinking wine.' (Currie 1995, 154)

Currie goes on to explain that if we didn't assign the same meanings to items in the imaginary context we could not keep track of the complicated connections between things that occur in such contexts. For example, the child pretends that there is water in her cup, then pretends to spill it, then wipe it up. If the pretend 'water' doesn't have the same semantic content as real water, the child wouldn't be able to make the connections between water being wet and spillable and something you can wipe up.

Next Currie considers the fact that the semantic degradation that Leslie describes doesn't apply to desire in the same way it applies to pretense. If metarepresenting involves decoupling then the same decoupling should occur with desire. I can pretend that there is water in the cup or I can desire that there be water in the cup. But the same deviance doesn't occur with desire. A child does not wipe up after an overturned cup about which he has the desire that it contain water. Currie concludes that whether the treatment of an expression is 'deviant'
depends, not on whether it is metarepresented, but on what attitude is taken toward the content. What may be deviant in the case of belief may not be deviant in the case of desire or imagining. On Leslie’s account, it seems that if one has a deficit in imagination, one should also have a deficit in the ability to desire. But this is not the case with children with autism.

Leslie’s theory was developed in part to explain the joint deficiencies in autism of lack of ability to ascribe intentional states and lack of ability to pretend. Both are due, according to Leslie’s theory, to an inability to metarepresent which involves decoupling. Leslie claims that belief attribution and pretend both involve decoupling because both involve referential opacity. The problem with this is that if referential opacity is the criterion for decoupling, then it is not just pretense situations and attributions of belief that exhibit this criterion. First order beliefs and desires also exhibit referential opacity (Currie 1995, 155). It’s true that if I pretend that a is f, and it is the case that a is identical with b, I do not thereby pretend that b is f. And it’s true that if I attribute to you the belief that a is f, and a is identical with b, I cannot assume that you believe that b is f. But, it is also the case that if I simply believe or desire that a is f, and a is identical with b, I do not necessarily believe or desire that b is f. If referential opacity determines when
something should be decoupled, Leslie's distinction between first and second order representations, where the second order is 'decoupled' from the first, is lost.

The metarepresentational/decoupling view thus cannot account for the joint deficiencies in autism. It would predict more than the two areas of deficiency. If one lacks the ability to metarepresent in imagination, then one should also experience the same problems in desiring and believing. But children with autism do not have these problems. In contrast, the simulation theory view is much more narrowly focused. As we have seen in our examination of Harris' view, a lack of ability to attribute mental states is due to a lack of ability of the imagination. The problem is not with the relationship of the representation to its normal semantic content, but of the relationship of the content to the attitude of the subject (Currie 1995, 156). That is, is the subject related to the content by way of belief, desire, or pretence? Children with autism don't seem to have a problem with metarepresenting, in Leslie's sense, except when it comes to the attitude of pretence. Furthermore, in addition to accounting for the deficiencies in imagination and mental state attributions, the simulation theory view presented by Harris offered an explanation of other autistic characteristics and tendencies. The simulation theory also explains facts about the lack of planning and problem solving abilities that were demonstrated in the 'Tower of Hanoi' problem. Let's now turn to
consider some further data on the abilities of children with autism.

The simulation theory hypothesis, it seems to me, provides a better account of why there is a variance in the understanding that is demonstrated by children with autism when it comes to mental states. As was discussed above, autistic children have no trouble with believing or wanting something to be the case. There are other examples where autistic children are successful in certain mind related tasks. Autistic children succeed in understanding some psychological states and not others. Studies have shown that children with autism have no difficulties, for example, in understanding that people may vary in their perceptual experiences. Peter Hobson concludes from his study, which compared normal children with children with a diagnosis of infantile autism, that autistic children are no more impaired than normal children in their recognition of visuospatial perspectives: "The autistic subjects were not markedly deficient in their ability to indicate the viewpoint of another 'subject' (in the form of a doll), nor did they fail to coordinate different viewpoints in the game of hide-and-seek" (Hobson 1984, 101). Autistic children also have no difficulty understanding that people may vary in simple desires and preferences (Baron-Cohen 1991; Harris 1990). But, as we know, autistic children have a very difficult time with false belief tasks. Almost all children with autism fail to understand the impact of second-
order beliefs, i.e., beliefs about beliefs (Baron-Cohen 1989; Ozonoff et al. 1991). These findings make sense if Harris is right and progress in understanding and ascribing mental states is achieved by means of an increasingly complex process of simulation. But they don’t make sense if your account doesn’t distinguish among mental states all of which are referentially opaque.

Children with autism do not show a general lack of ability to theorize. Analogue tests to the false belief test were given to children with autism. These tests involved false maps and false photographs. Let’s first review the false belief task (Baron-Cohen, Leslie, and Frith 1985). The children are told the following story: Sally has a basket and Anne has a box. Sally has a marble which she puts in her basket. She then goes out. While Sally is away, Anne takes out Sally’s marble and puts it into her box. Now Sally comes back and wants to play with her marble. Where will she look?

On the photograph test Sally is replaced by a camera and Sally’s belief is replaced by a photograph of the marble in the basket. The photograph is placed face down so the child can’t see it. The marble is moved from the basket to the box and the child is asked where in the photograph the marble is located. People with autism actually perform better than normal four and a half year-olds on this and the false map task (Leslie and Thaiss 1992). In
addition, when inferential reasoning is required to make judgements in stories involving mechanical or behavioural rather than mental events, people with autism again perform better than normal four and a half year-olds. These results are explainable both on Leslie's theory theory view and the simulation theory. The reason, however, that they are explainable on Leslie's view is not because it is a theory theory view, but because it is a modular view.

It is obvious how the simulation theory would account for the discrepancy in theorizing abilities. The simulation theorist would claim that since the lack of ability of autistic children to attribute mental states has nothing to do with theorizing, a lack of deficiency in theorizing in non-mental areas is not surprising. In fact, it is something that the simulation theory would predict. The theory theory, on the other hand, cannot explain this discrepancy unless it is also a modular version of the theory theory. If one, for example, takes the problem with autism to be some kind of central processing problem (e.g., Frith 1989), it is not clear why the processing would fail in this one domain and not others. Because Leslie holds a modular view, where what is missing is theory of mind information, he can also account for the lack of deficiency in these other areas. In fact, he uses it to support his modular view. But the fact that children with autism are selective in their theoretical failure does not, in itself, imply that what they
are missing is a theory in that domain. All this data tells us is that if you hold a theory theory view, it must be modular. It does nothing to support the theory theory view in general.

There are two other characteristics of interest that are prevalent in children with autism. Autistic children have a preference for narrowly focused, fact-based topics and they are averse to any change in their environment or disruption of their routines. These characteristics, like the lack of difficulty in theorizing, find obvious explanations if the simulation theory is correct. The fact that we find the topics pursued by children with autism to be tedious and boring may simply be a result of the fact that we have imaginative abilities that they lack: "If autism is, at bottom, a failure of imagination, the activities most of us find interesting are simply not available to autistic people" (Currie 1996, 255). And if people with autism cannot imagine future or hypothetical situations, it is natural that they would prefer routine and consistency in their environments. A change would signal something frightening not because it unknown but because it is unimaginable.

Lorna Wing and Judith Gould (1979) have discovered the presence of a triad of impairments in children who have been diagnosed as classically autistic. The triad includes impairments of social

37The 'classic' autism label is contrasted with other labels taken to identify different forms of autism such as Asperger's Syndrome or fragile X syndrome. Not all children with these latter
competence, communication skills and pretending. As we have discussed, the simulation theory appears to offer a more natural and defensible explanation of the impairment in pretending. The theory theory, however, can explain the other two areas of impairment just as well as the simulation theory. Since it’s clear that children with autism have trouble with mental concepts, it is this difficulty that explains their impairments in communication and social skills. Without the ability to ‘read between the lines’ when someone is speaking or see facial expressions or other behaviour as indications of underlying feelings or thought, communicating and socializing, as we understand them, are virtually impossible. The autistic person simply does not understand what is going on. Uta Frith explains:

The same autistic child who can understand very well why a customer pays the shopkeeper, or why a person jumps out of the way of a falling rock, may not understand why a polite guest declines a further offer of food when he is still hungry; why an employee who wishes to be promoted gives flowers to the boss’s secretary; why a schoolgirl complains of a stomach ache whenever she has not done her homework; or why a toddler exaggerates his hurt by crying when his brother pushed him. (1989, p.176)

Both the simulation theory and theory theory explain the social and communicative impairments of the basis of a lack of understanding of intentional states. The distinguishing question then will not be related to explanations of these impairments diagnoses will display autistic characteristics.
but will be which theory better explains the lack of understanding of intentional states and which explanation also works to explain other characteristics associated with autism. I believe that the simulation theory is the more natural and inclusive theory in this regard.

I have argued in this thesis that realism about intentional states is not jeopardized by externalist arguments. If the ability to simulate is at the root of our ascriptions of mental states to others then we have a means of identifying the existence of mental states that is independent of our context-bound means of individuating them. While I don't take myself to have definitively demonstrated the truth of the simulation theory, I do believe that I have demonstrated that it appears to have much to recommend it and, that being the case, we don't have to worry about defending psychological realism anymore.
REFERENCES


218

