On Internalist Intuitions
A Study in the Externalist-Internalist Debate

A Dissertation Submitted
for the degree of M. Phil in Philosophy
2001

by

Madhucchanda Sen
Department of Philosophy, University College London
ABSTRACT

On Internalist Intuitions

A Study in the Externalist-Internalist Debate

A Dissertation Submitted for the degree of M.Phil in Philosophy

2001

Madhucchanda Sen

Department of Philosophy, University College London

This thesis is on the externalist-internalist debate in the Philosophy of Mind. Externalism and internalism are two contesting theses about the individuation of mental states. Externalism is the thesis that there are some mental states which entail the existence of the objects on which they are directed, or that the thinker's environment is a certain way. Internalism, on the other hand, is the thesis which says that mental states of persons are such that there is 'no necessary or deep individuative relation' (Burge) between the individual's being in states of those kinds and the nature of the individual's physical and social environments. Consequently no mental states can entail the existence of the objects which the states are directed towards, or entail that the environment is a certain way.

One major worry associated with the debate is that it is hard to find common ground between internalists and externalists in terms of which we can make sense of the debate. The Twin Earth stories were originally used to generate intuitions in support of externalism but these intuitions are controversial. I propose that the contending parties have other commitments which explain why they take stands they do. Those other commitments are the ones to be questioned and hence the solution
to or the resolution of the debate can be attained only by questioning those commitments. In this thesis I talk about two such presuppositions that the internalists have made. I give a defence of externalism by questioning those presuppositions.

The first chapter of the thesis is a survey of the debate and an investigation of the motivations that lead the contending parties to uphold their positions.

The kind of support that I give to externalism is a negative one. It comes in form of an attack on the following presuppositions of the internalists: (a) We can consider two mental states to be the same (i.e. type identical) if the subject cannot discriminate between them; and (b) We can consider two mental states to be the same if they give rise to the same behaviour.

The second chapter is an attack on the first presupposition. It is argued that there are instances of mental states where subjective indiscriminability cannot be regarded as evidence for sameness. The third chapter counters the second presupposition by showing that the notion of behaviour that the internalists are invoking in order to hold onto this presupposition is a very restricted notion of behaviour. However, externalism seems to face the problem of mental causation, so accordingly, the fourth chapter deals with this problem. Davidson has a clear answer to the problem of mental causation and he is also an externalist. And so there is a clear defence of the compatibility of mental causation and externalism, which I endorse.
## Contents

<table>
<thead>
<tr>
<th>Contents</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td>6</td>
</tr>
<tr>
<td>Chapter One</td>
<td>9</td>
</tr>
<tr>
<td>1.1 Introduction to the Debate</td>
<td>9</td>
</tr>
<tr>
<td>1.2 The Origin of the debate</td>
<td>13</td>
</tr>
<tr>
<td>1.2.1 Hilary Putnam and his Thought Experiments</td>
<td>13</td>
</tr>
<tr>
<td>1.2.2 Tyler Burge and his Thought Experiments</td>
<td>21</td>
</tr>
<tr>
<td>1.3 Internalist Responses</td>
<td>26</td>
</tr>
<tr>
<td>1.4 Marr's Theory of Vision and the Externalist-Internalist Debate</td>
<td>41</td>
</tr>
<tr>
<td>1.5 A Puzzling Situation</td>
<td>48</td>
</tr>
<tr>
<td>Chapter Two</td>
<td>54</td>
</tr>
<tr>
<td>The Identity(-Individuation) Crisis</td>
<td></td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>54</td>
</tr>
<tr>
<td>2.2 What it all hinges upon: Cartesianism and Internalism</td>
<td>58</td>
</tr>
<tr>
<td>2.3 Identity and Indiscriminability</td>
<td>84</td>
</tr>
<tr>
<td>2.3.1 General Difficulties</td>
<td>87</td>
</tr>
<tr>
<td>2.3.2 Specific Difficulties</td>
<td>91</td>
</tr>
<tr>
<td>2.4 The Infallibility Thesis</td>
<td>92</td>
</tr>
<tr>
<td>2.5 The Problems with Phenomenal Character</td>
<td>101</td>
</tr>
<tr>
<td>2.5.1 The Paradoxes of Phenomenal Character</td>
<td>103</td>
</tr>
<tr>
<td>2.5.2 Problems with Qualia</td>
<td>105</td>
</tr>
<tr>
<td>2.6 The Common Element View: Philosophy of Mind and Philosophy of Perception</td>
<td>115</td>
</tr>
<tr>
<td>2.7 Externalism and Self-Knowledge</td>
<td>124</td>
</tr>
<tr>
<td>2.8 Subjective Indiscriminability as Good Evidence</td>
<td>138</td>
</tr>
<tr>
<td>2.9 What we can conclude</td>
<td>140</td>
</tr>
</tbody>
</table>
Preface

My research on the Externalist-Internalist debate started in Calcutta in the year 1994-1995. I was then a Junior research Fellow of the Indian Council of Philosophical Research attached to the Department of Philosophy in Jadavpur University, Calcutta. Later on I moved to the Department of Philosophy at University College London with a Commonwealth Scholarship. My views on externalism and the externalism/internalism debate were different from what they are now. The way in which I see the debate now, is mostly due to the London academic exposure. The orientation of my research on this issue in Calcutta had been from the perspective of De Re thoughts. It has now been replaced by the perspective of twin-earth thought experiments and problems relating to mental causation. I would like to thank all those people and institutions who are responsible for this change in my perspective on the issue. Of course the Commonwealth Scholarship Commission and the British Council come first on the list. I would like to thank the Commission and the Council for giving me this opportunity to study in UCL and for all their help during my stay in London. Special thanks are due to the Department of Philosophy at UCL for their academic and financial assistance.

Much of my views on this issue have been questioned and later on influenced by my supervisor Dr. Tim Crane. I consider myself to be extremely fortunate to have had Dr. Crane as my teacher and supervisor, whose unparalleled interest in the well-being of his students, and unfailing support and encouragement has made this dissertation possible. I would like to express my deepest respect and gratitude to him.
I have benefited greatly by the guidance and comments of my secondary supervisor Dr. Michael Martin, especially on the issue of identity and indiscrimination. I would like to thank him for his criticisms and bibliographical help, and especially for providing me with a copy of his paper ‘The Reality of Appearance’ which has shaped much of the views I have expressed in the Second Chapter.

Much of my views on this debate took shape after attending a joint seminar on Content and Externalism given by David Papineau and Gabriel Segal in Kings College, London in 1995-1996. I would like to express my thanks to them for inspiring me by their lectures.

I take this opportunity to thank Paul Noordhof for his detailed comments on my thesis.

Going back to Calcutta, I would like to thank the Department of Philosophy, Jadavpur University, and my teachers there. Especially to my two previous supervisors, Dr. Amita Chatterjee and Dr. Chhanda Gupta from whom I have learnt so much and who have always shown ceaseless interest in my work. I have been greatly helped by Dr. Chatterjee’s comments on my work.

Back in England my two academic friends Lucie Antoniol and Giovanna Hendel have been extremely helpful. They have always patiently listened to my tormented ramblings and I have benefited greatly from their invaluable remarks. Sophie Allen, my fellow student at the department, has been singularly helpful with her comments on the debate. I thank her too. My other friends have played a big role in my
London struggle. Special thanks are due to Rinaldo Brau, Itziar Ochotorena, and Francesca Fabri for their computer assistance. I would like to thank Alessandro De Roma, Arnaud Roitg, Claudia Cohen Añorve, Christine Ling, Francesca Flamini, Iñigo Ballester, Itziar Ochotorena, Luis Diaz, Matteo Lippi Bruni and Nuria Sandino who have been constant source of strength and support. Without their love and care it would have been really difficult.

My greatest debt is surely to my family. To my parents who have given me everything. I thank my mother, Mrs. Rama Sen for her encouragement and support and for passing on her love for academics, to me. My debt to my father and teacher Professor Pranab Kumar Sen is undoubtedly the greatest. He gave me love and support and also passed on the greatest love of his life to his children: Philosophy. I have been singularly fortunate to have had an excellent teacher and philosopher as my father, but have been equally unfortunate to have lost him just in the middle of my longest and hardest academic struggle. Much of my views on propositional attitudes have formed after continuous discussion with him and my sister Manidipa, and I thank her too. Finally, I would like to thank my husband, Nirmalya Chakraborty for everything, and especially for being uncompromisingly insistent on my being respectful towards the duties I have to myself as an academician.
Chapter One


1.1 Introduction to the Debate

Externalism and internalism¹ in the philosophy of mind are two conflicting theses about how intentional mental states² are correctly individuated. Externalism about the mind is the thesis which says that there is a deep individuative relation between an individual’s being in those states and the nature of her social and physical environment. The very existence of those states entails the existence and the nature of the objects towards which they are directed. Internalism about the mental, on the other hand, is the thesis that there is no such deep individuative relation between the individual’s being in states of those kinds and the nature of her physical and social environment. Externalists have tried to give us instances of intentional states which are such that the thinker could not be in such states unless her external environmental circumstances were the way they were. And so to have a full understanding of such states or to individuate such states (i.e. identify them and to distinguish them from others) it would not be sufficient to know what goes inside the head of the thinker alone. Rather, we shall have to look into the external environmental causes which determine what goes inside the thinker’s head. The Putnam-kind of ‘Twin Earth’³

---

¹ Though some philosophers prefer to call this view ‘individualism’, I have called it internalism throughout in this thesis. Obviously the two terms mean the same for me.
² I am not taking any stand on the issue regarding whether all mental states are intentional or not. But I am going to discuss only the intentional mental states, as it is in the case of the intentional mental states that the externalist-internalist debate arises. We will however speak about other states in the Second Chapter.
thought experiments were designed to show this, which I shall describe below. Internalists, however, draw a different lesson from the very thought experiments which the externalists had used to prove their thesis. They (internalists) suggest that what the externalists were missing was the significance of the fact that the world appears to the two protagonists in the thought experiments (which will be discussed shortly) in exactly the same way. If they were to switch places, they would not notice it at all. And so the internalists conclude that mental states are individuated by their narrow, internalistic properties, as opposed to broad/wide, externalistic properties (in other words, intrinsic non-relational properties as opposed to extrinsic relational properties).

The question regarding how we ought to individuate mental states is at the centre of our concern about the mind and the mental. Following Quine's dictum 'No entity without identity', we all are in the quest of a criterion of identity for any entity we want to include in our ontology. If we are to include mental states (intentional mental states) in our ontology, then we need to have an identity criterion for such states. And in order to have such an identity criterion, we have to say what the conditions of sameness and difference of mental states are. And it is here that the externalist-internalist debate comes in. It is a debate about how one ought to individuate mental states.

But why should we, philosophers, want to include mental states in our ontology, at all? What are the considerations that lead us to believe in the existence of such mental states, in the first place? In finding a suitable criterion of identity of mental states, we have to keep in mind those considerations which had led us to believe in the existence of such
states in the first place. If we forget those considerations, while construing the criterion of identity of mental states, then we might end up with entities which do not serve those purposes which mental states seemed to serve and hence were felt as necessary parts of our ontology. So what should be our prime consideration when we are trying to individuate mental states? The notion of individuation has been used in the philosophy of mind as one closely linked with the notion of explanation. Usually mental states are introduced in philosophical discussions, because they are seen to be figuring in explanations of actions. For example, we speak about beliefs and desires since they figure in explanations of actions, which are taken to be caused by those beliefs and desires. The internalists maintain that the kind of explanations that we want to give in the philosophy of mind can be given by individuating mental states purely internally. On the other hand, the externalists suggest that in order to give psychological explanations we have to presuppose the existence of the objects towards which they are directed.

We find in the literature that this debate is often conceived as a debate about what mentalistic content should be like: broad/wide or narrow. This is associated with the view that if we have to identify a mental state, we have to identify it in terms of its content. Often a belief is identified and differentiated from other beliefs in terms of its content. So how we conceive of content of mental states is the key to how we conceive of our mental states. The debate is also conceived in terms of a debate about the narrow and wide contents of thoughts. What is meant by wide and narrow content? Content externalism claims that the contents of states are dependent on external factors. But this way of construing the externalist thesis is not useful since an internalist could
agree that the content of a mental state depends on many such things. And so for them too content would be wide in a trivial sense. For example, the content of the sentence ‘The present King of France is bald’ entails the existence of the King of France (or presupposes, if you like, the existence of the king). But the real point is whether or not having a thought with that content would entail the existence of it.

The debate as we find it today centres around these questions: 1. Questions regarding the role of mental states in psychological explanations in the philosophy of mind; 2. The question regarding how contents are to be conceived, if mental states are to figure in psychological explanations by virtue of having those contents. Or even more fundamental questions like: 3. What kind of criterion of identity should mental states have? And, 4. Do we need to individuate mental states in terms of their contents in order to make them figure in psychological explanations at all? In this thesis we shall see how the externalist-internalist debate deals with all these questions. I aim at coming up with a negative defence of externalism by showing that the stands that the internalists take up in responding to these questions might not be acceptable altogether. And in doing so I shall be saying that the internalists could be said to be moving with two assumptions which have led them to take these stands which might not be acceptable. In the Second Chapter I will be discussing what could be extracted from the internalist’s position in the debate as their accepted criterion of identity of mental states. In the Third Chapter I shall discuss another view of the internalists that might be extracted from the debate which determines their response to the first two questions listed above. In the Fourth Chapter I discuss a very important question which is closely linked with
the fourth question listed above, viz.: How can we place the mind and the mental in the physical world of causes and effects?

This chapter is an introductory one. Here I discuss the debate as it is found in the literature. And it is here that I try to set the stage for the battle to be fought by the contending parties on the questions that I have listed above.

1.2 The Origin of the debate

Hilary Putnam is responsible for the birth of the debate. His views about how reference is achieved have led to the externalism-internalism debate. The various thought experiments he offers to prove his thesis about reference can be extended to the philosophy of mind to have support for an externalist thesis. Though such thought experiments were originally used in the philosophy of language, they had a natural extension to philosophy of mind. I shall try to state here the thought experiments that we find in Hilary Putnam and Tyler Burge and try to bring out the conclusions they draw from these experiments.

1.2.1 Hilary Putnam and his thought experiments

In Putnam we find a host of thought experiments which were originally constructed to prove or clarify his position on issues central to the philosophy of language, especially to the problem of reference. We actually get two varieties of thought experiments in Putnam’s work,

though both the kinds aim at proving the same thing. His general dictum about meanings is that they are not ‘in the head’. And we find him coming up with thought experiments of the following kinds: 1. The first kind tries to show that I might be exactly similar to someone, as far as what goes on in my head is concerned, yet while my words may mean/refer to something, her words may mean/refer to something completely different, just because our environments are different. 2. The second kind tries to show that unless the subject has an external causal link with the object, the name or singular term that she uses for the object cannot succeed in referring to it.

Applied to the realm of thoughts, these experiments show that certain thoughts are such that we could not have them unless the objects to which they are directed did not exist. So the very existence of the thoughts entails the existence of the objects that they are about.

The thought experiment that I shall concentrate on here is of the first kind and is known as the ‘Twin Earth Argument’. Before going into a discussion on it I want to point out one thing, viz., in understanding and in assessing this argument we should never lose sight of the original motivations that Putnam had in giving his thought experiments and the arguments resulting from them. In ‘The Meaning of “Meaning”’ we find Putnam trying to criticise a particular view about reference and in *Reason Truth and History* he is trying to show how reference-success can be achieved. In the first case he comes up with his twin-earth thought experiment and in the second he gives us his famous Brain-in-a-Vat

---

argument. I think that the Brain-in-a-Vat argument is also important for us, and I shall explain why I say this later on.

Now onto the thought experiment. To start with, we find Putnam speaking about two presuppositions which traditional theories of meaning (which he challenges) hold.

1. 'knowing the meaning of a term is just a matter of being in a certain psychological state'; and,

2. 'the meaning of a term determines its extension (in the sense that the sameness of intention entails sameness of extension).

Putnam suggests that no theory of meaning can entertain both of these presuppositions together. And this is brought out by the thought experiments and they make it clear that the two cannot be adhered to together and we must drop one of them.

The thought experiments demonstrate that meanings are not in the head in the sense that knowing a meaning is not just being in a psychological/mental state which is strictly inside the head (i.e. a mental state with a narrow content). This is so because a psychological state with narrow content cannot determine the meaning which can determine extension. This is what the twin earth thought experiments have allegedly demonstrated. It is only if one is in a wide psychological state, that we can say that her psychological state can determine the meanings of her words. And this meaning in its turn determines the extension, even in the cases spoken of in the twin earth thought experiments. And so meanings are not in the head and mental states are not narrow, if meanings have to determine extension.

---

We shall see how the thought experiments demonstrate this. Putnam says that if we accepted both the presuppositions which traditional theories of meaning hold, then we would have to say that the sameness of internal psychological states entail the sameness of meaning. In the case of the twins in the twin earth thought experiments this is shown to fail. We shall now see how.

Putnam urges us to imagine of a far away planet infested by doppelgängers of each of us on earth who are atom-to-atom identical with us. Let us call this planet 'twin earth'. A section of inhabitants of this planet are like their English-speaking counterparts. The planet too is very much like our own planet. The only difference that this planet has with earth is that the thing that the twin-earthers call 'water' is not H₂O, but is something else, say XYZ, though it has all the other observable properties of water in earth. As a result, when our doppelgängers talk about XYZ, they have exactly the same thing going on in their heads as what goes on in our heads when we talk about H₂O. What is important is that, while they refer to XYZ by the use of the word 'water', we refer to H₂O, by using the same word 'water'. Consequently, in spite of being in the same psychological state, the meaning of 'water' in the twin-earther's mouth and the meaning of 'water' in our mouth are different as they refer to two different things. Suppose an educated chemist from earth were to go to twin earth, then she would say:

'On twin earth the word "water" means XYZ'.

Similarly if a chemist from twin earth were to visit earth, then she would say:

'On (twin) earth the word "water" means H₂O'.

Putnam asks us to think of a time when both the inhabitants of earth and twin earth had almost no knowledge of chemistry. In such a situation neither Ed from our planet, nor his doppelgänger Ted (the twin Ed and hence T-Ed or Ted), from twin earth, would be able to detect that his twin was referring to a completely different thing by the use of the same innocent term ‘water’. But surely could this suggest that the word ‘water’ had two extensions? By the word ‘water’ Ed meant $\text{H}_2\text{O}$ while Ted meant $\text{XYZ}$, although they were literally in the same internal psychological state. So it seems that you cannot have both the presuppositions. If meanings were to determine extensions, then we have to say that in the case of Ed and Ted there must have been some difference in meaning which had led to the difference of the extension. And if the meanings were different in spite of the fact that Ed and Ted were in the exactly identical narrow internal mental states (in other words, in spite of the fact that what was strictly inside their heads were identical), then meanings could not be in the head. Hence Putnam concludes: ‘Cut the pie any way you like, “meanings” just ain’t in the head’. So we see that the thoughts of the twins have to be individuated externalistically (keeping in mind the differences in their environment) in order to account for the obvious difference in the extension of their words. Thus Putnam gives us an argument in form of a thought experiment for Externalism.

As I have said before, this and the other thought experiments were offered by Putnam just to show how reference-success was possible. One may feel tempted to bring a sceptical argument against Putnam which tries to show that we cannot refer to anything at all. The argument is a typical one that might be brought forward by a global

---

7 ‘Meaning and Reference’, p.155.
sceptic. It might run along the following line: If Ted does not manage to refer to water in twin-earth by uttering ‘water’, and given that what goes inside his head is the same as what goes inside the head of Ed, then how can we be sure that even Ed manages to refer to water? Maybe Ed is suffering from a global delusion; what he thinks is water is not water at all. But against such an argument we have Putnam’s Brain-in-a-Vat (BIV, from now on) argument. An argument which shows that the supposition that we are BIV’s who can never refer to things outside is really self-refuting. This is so because, in a nut-shell, a brain in a vat cannot either be right in thinking/saying that it is a brain in a vat, or in thinking/saying that it isn’t. And so there is no way of formulating the position.⁸

The BIV situation is actually analogous to the Cartesian demon situation. Suppose that my brain was taken out of my cranium and was kept alive in a vat of nutrients. Suppose further that my brain was connected with a super-computer which managed to send in and send out stimuli in such a way that the world seems to me as though it had not been changed at all. In such a situation I might be having thoughts and might even be speaking (or having speech experience from the inside) and yet whatever I speak would not refer to anything and whatever I think would not be about the objects that I think they are about. Seen from the inside what I feel in this situation is exactly the same as what I feel in the normal situation. But do my words refer to anything, in the vat case? They do but only to vat-‘objects’ in the vat-‘world’ and not do not refer to real objects in the real world. And so the sceptics conclude that since seen from the inside the two situations are the same, if I cannot refer in one I fail to do it in the other too. As we

⁸ See Putnam’s *Reason Truth and History*, pp. 5-21.
have already mentioned, Putnam says that there is something inherently wrong in this argument as it is self refuting. This is so because the BIV can neither be correct in saying that it is a BIV nor be correct in saying that it is not a BIV. This is so because if it is indeed a BIV then its words are in vat language and not in natural language and can refer to vat objects in the vat world and cannot refer to real objects in the real world. But it is in the real world alone that it is a BIV. So its utterance 'I am a brain in a vat' will be true if and only if it is false.

Putnam does not stop with this negative argument against the sceptical position; he goes on to give a positive thesis about how exactly referential success is achieved. On the one hand, he gives us a transcendental argument showing that we do succeed in referring, on the other hand, with the thesis on 'division of linguistic labour' he works out the socio-linguistic phenomenon that is behind such a success. The division of linguistic labour thesis is that in a linguistic community we see a kind of division of linguistic labour. There are at least certain terms that are used by a linguistic community, the associated criteria of application of which is only known to a subset of the community, viz. the experts. However this knowledge is transmitted to the ordinary users of the language from the experts.10

What Putnam wants us to realise is that though not all of us have the technical knowledge which enables us, for instance, to distinguish aluminium from molybdenum, or elms from beeches, but it is not as though no one possessed or could possess it. The linguistic community is

---

10 See Putnam's 'Meaning of Meaning.'
'considered as a collective body', a body which divides the labour of knowing and employing such words.

This brings out more clearly why the two classical presuppositions cannot be maintained together. And it also shows that our narrow, strictly internal psychological states (which may tend to be the same for elms and beeches) do not in the least fix the reference.

This is Putnam's externalistic argument. We have to remember that the externalist, following Putnam, is not saying that 'water' has a constant world-invariant meaning, i.e. 'water' means the same in earth and twin earth, but water in earth is H₂O while it is XYZ in twin earth. What she wants to say is that water is H₂O in both earth and twin earth and if you like, in all possible worlds, it is just that the term 'water' in twin earth has a different meaning. What Putnam intends is that if we discover that water (in the actual world) is H₂O, then it will be H₂O in all possible worlds.

Thus meanings cannot at least always be just what are determined by what we have in our head and our intentional mental states are not individuated by what we have in our head; they can be causally determined by external factors. If it were the contrary, then there would have been no difference between me and the brain in a vat; but we have seen how that cannot be the case. And Putnam has not just stopped at showing that the BIV argument is self-refuting, but goes on to show how the utterance-object link is established. This is clear from Putnam's division of linguistic labour thesis. Here he speaks of the expert who acts
as an intermediary that links us to elms and beaches and to aluminium and molybdenum.

Just as the meanings of Ed and Ted’s ‘water’ utterances are not in their heads, the individuation of Ed and Ted’s thoughts (in this case ‘water’-thoughts) cannot be made purely by referring to what goes strictly inside their heads. Ed and Ted might have the same thing going inside their head, just as they share all their physical properties, and yet Ed’s thoughts are water thoughts whereas Ted’s are at best twater (twin-water) thoughts. There must be some way of showing that their thoughts are different, otherwise we would not be able to explain why Ed’s thoughts (which look from the inside same as that of Ted’s) result in H₂O involving behaviour, while Ted’s thoughts result in XYZ involving behaviour. Analogously, if I am reduced to a BIV, then my thoughts would never be genuine thoughts in spite of the fact that internally they seem to me the same as the ones that I had when I was not a BIV. This is how Putnam’s thought experiments carry over to the realm of thoughts.

1.2.2 Tyler Burge and his Thought Experiments

We find Tyler Burge countering internalism in 'Individualism and the Mental' and 'Individualism and Psychology'. In these essays we find him coming up with Putnam-type thought experiments which are used against internalism. (We should, however, keep in mind that Burge is an internalist about indexical and demonstrative singular thought.)
The arguments that we find in the first two papers are more relevant for our purpose, as in these two papers we find Burge giving us reasons for adopting an externalistic individuation of mental states. The reason that he gives is closely tied with the question of what the nature of psychological explanation should be. The question 'Is the externalist characterisation of mental states better suited for explanations in psychology than the internalist one?' is a central question for this thesis. The internalist theses I intend to discuss here in detail are theses which try to establish that a narrow, internalistic individuation of mental states is best suited for psychology. And so 'Individualism and the Mental' and 'Individualism and Psychology' are more important for our main discussion in this thesis. However the other two essays too will throw light on the debate.

There is one issue that has to be mentioned here. Whenever we are talking about mental states and how they should be individuated, we have to see if the way that we are individuating mental states makes them suitable for explanations in psychology. This is actually because each of the contending parties (viz. the externalists and the internalists) claim that the way they conceive of mental states is the way that is best suited for explanations in psychology. But we have to remember that there are two different questions that we have here, because by psychology we can either mean scientific psychology or folk psychology. So when we are asking the question 'Is the externalist characterisation of mental states better suited for explanations in psychology than the internalist one?', we actually could be asking this question about folk psychology or about scientific psychology. Which of the two different questions are important to us would be a matter dependent on which of the two psychologies we regard as important. Now maybe not all
philosophers are keen on having a science of the mind and this is may be
due to the scepticism regarding the success of such a science. However
the distinction should be made. I, however, feel that what we say in
scientific psychology should not be totally different from what we
commonly hold as intuitive. I shall not be directly concerned with the
debate between folk psychology and scientific psychology in this thesis.
What I want to emphasise is that what kind of psychology we are going
to opt for is very often determined by what notion of mental states we
are working with. On the other hand, what kind of mental states we are
concerned with depends upon what kind of psychology we are opting
for. So I find it difficult to decide what kind of psychology we need to
have without deciding what is it that we need to explain and I find it
equally difficult to decide what is it that we need to explain without
deciding what kind of psychology we can have. I find it even more
difficult in deciding what kind of psychology we need to have because
the distinction between folk and scientific psychology is not always quite
clear in the literature itself. I hope that the distinction will be clear as we
proceed with the discussion. In the third chapter we shall specifically
discuss this.

In his paper 'Individualism and Psychology', Burge criticises a
few arguments offered by internalists philosophers in favour of their
thesis that explanation in (scientific) psychology is and ought to be
purely 'individualistic' (as Burge calls it). He then offers us an alternative
psychological theory which is not individualistic, yet which he claims to
be equally powerful. In the end he offers us an argument against
individualism, an argument which centres around visual perception.
Burge says\textsuperscript{11} that the recent prominent versions of internalism, as we find in philosophy of psychology, can take two forms. The first form maintains that the fact of one's being in a particular mental or intentional state can be explicated purely and wholly by reference to states and events of the individual without having to bring in or having to presuppose anything whatsoever about the individual's socio-physical environment. The explication is done purely in terms of stimulations, behaviours, internal physical and functional states of the individual. For example, in functionalism it is regarded that what makes the mental states the type of state it is, is the functional relation that it has with the subject's stimuli and behavioural responses and to other mental states.

The other form of internalism is not aiming at an explication of mental states directly. It makes a claim of \textit{supervenience}. The claim is that the fact that a person has a particular intentional state is possible only because the individual has particular physical, chemical, neural and functional states, and such states could be specified without ever bringing in the socio-physical environment of the individual. This, I feel, is probably the strongest argument that the internalists have against the externalists. Internalists claim that if we want to make sense of how the mind and the mental can causally interact with and in the physical world, then we have to respect this supervenience. But the thought experiments of the externalists and the conclusions that are derived from them seem to violate this supervenience. Violation of supervenience which comes along with the broad conception of content would result in serious problems in the question of causation. And hence their slogan: If we have to match content with causation, content has to be narrow.

rather than broad. I shall be discussing this line of internalist attack in detail in the third and the fourth chapter.

I mention the supervenience thesis here particularly because, Burge has argued in a few of his papers that a person might have exactly identical physical and functional histories as another, and yet the two can never be said to be in the same intentional/mental state. In fact the thought experiments that Burge comes up with are showing exactly this. And so the Burgian thesis is a direct denial of the supervenience claim.

Burge talks of two persons, A and B. Suppose A is capable of identifying aluminium. She has the following beliefs about aluminium: it is a light metal and it is used in sailboat masts. A, however, is ignorant about the chemical structure and micro-properties of aluminium. Imagine a counterfactual situation where A retains all her internal histories but the environment she is in lacks aluminium. There is however in this counterfactual environment a metal very similar to aluminium and is also called 'aluminium'. In such a situation it is quite obvious, according to Burge, that A would lack an aluminium thought though her internal histories are all the same in this environment.

Let us take up the other thought experiment that Burge has given. Suppose again that B believes that she has arthritis in the thigh. B however can identify arthritis and knows facts about it. She is just ignorant of the fact that one can never have arthritis in the thigh, it is a disease which cannot occur outside joints. Now imagine an environment

\footnote{See his 'Individualism and the Mental, 'Other Bodies', and 'Two Thought Experiments Reviewed' (in \textit{Notre Dame Journal of Formal Logic} 23, 1982).}
where no one has ever isolated arthritis as a disease only occurring in the joints, and the thing that B has in her thighs is also called 'arthritis'. Burge argues that B would surely, in such a situation, lack any arthritis thought, though her internal histories are the same in this counterfactual situation. This difference in A and B can be accounted for only if we bring in the changes in the environment.

What these arguments are out to prove is that mental states and events can in principle vary because of variation in the environment, even though the internal functional and physical histories of the individual remains absolutely the same. Consequently the individuation of the mental states in terms of the functional, internal histories specified non-intentionally cannot be of much help.

1.3 Internalist Responses

Maybe the most general way in which the internalists have responded to the externalists is that the externalist's characterisation of mental states doesn't help us to have the kind of causal explanations that we need in psychology. I am not going to start with this general response of the internalists. This will come out as we proceed our discussion. I would like to start off with the way that internalists have responded to the thought experiments of Burge and Putnam and then go on to show what could have been the internalist assumptions that might have led them to react in this way to the externalists. Now there could be two different approaches to the thought experiments and their use in favour of the externalist story. On the one hand, one could take these thought experiments as serious threats to internalism and try to counter them. On the other hand, one might just take these to be posing no threat at all to
the internalist's thesis. I shall start off with the discussion of the view that the twin earth thought experiments do not pose any serious threat to the internalist thesis.

Tim Crane in his 'All the Difference in the World' has questioned the twin earth intuition itself. Crane has reservations about the significance of the twin earth thought experiments in general and also has specific reservation against Burge's linguistic thought experiment. To start with, Crane points out that the arguments that the externalists try to construct from the twin earth thought experiments undermine a very important point about causation. They fail to appreciate that it is only our intrinsic properties that can be causally efficacious. What internalists, like Crane, want us to realise is that, in order to make mental states causally efficacious we cannot individuate them widely. If we are to follow the twin earth intuitions and think that the broad contents of our beliefs can cause our behaviour (and we could make behavioural predictions about a person if we could know the broad contents of her beliefs), then the broad content has to be intrinsic properties of ours. But our being in certain external circumstances can never be regarded as our intrinsic properties. Consequently, Crane says, it is difficult to see how broad content can fit into the causal story. There may however linger a few questions about Crane's argument against the twin earth cases in general. What would count as intrinsic property? Why is it that only the intrinsic properties that can be causally efficacious.

---

14 We shall discuss this issue in greater details in the fourth chapter.
efficacious? And of course along with this the whole issue of mental causation opens up.\textsuperscript{15}

Crane has a specific objection against Burge's linguistic thought experiment. Let us consider the arthritis example. In the example, the person concerned, i.e. \( B \) in the actual situation is claimed to have a false belief. Crane asks us to look closely into what could have been the source of such an erroneous belief. We have to remember that when \( B \) says that he has arthritis in his thighs, he does not only have a belief about arthritis or his physical condition, he also has a belief that such a belief could be expressed by the sentence 'I have arthritis in my thigh'. This is true of any utterance made by a normal language user whatsoever. If I say 'I love to go and sit in the Royal Festival Hall foyer and drink coffee every now and then', then I do not merely have a belief about my love for Royal Festival Hall or about what I enjoy doing on a lazy day in London. I also have to have a belief about which words in the English language can express this belief of mine. Suppose I express the same belief in Italian by saying 'Mi piacerebbe andare di tanto in tanto a prendere un caffé nel foyer del Royal Festival Hall'. In this case I have associated with my beliefs about the Royal Festival Hall and coffee and what I enjoy, also my beliefs about the Italian words and the sentence which best capture this belief of mine. And when we are suggesting that there has been a mistake on \( B \)'s part, then we have to find out which of the two beliefs went wrong. \( B \) could be mistaken about his beliefs about arthritis or he could be mistaken about which words he should use to express his belief, or both. \( B \) might well have had a public arthritis concept and so might well have had a false belief when he thinks it to be

\textsuperscript{15} This I shall discuss in the fourth chapter. And following a Davidsonian line, we shall see how even the external can count as intrinsic in the discussion of self knowledge in
affecting his thighs. He could even just have a false belief about which words he ought to use when he speaks of a disease which affects his thighs. Alternatively, B might not have a proper arthritis concept at all, and have, what Burge calls, a *tharthritis* concept which applies to both arthritis and whatever disease affects his thighs. In such a case, B’s first level belief to the effect *I have tharthritis in my thighs* is true but he is mistaken when he thinks that the English term that expresses his *tharthritis* concept is ‘arthritis’.

Now which of the two ascriptions would be correct? There seems to be no linguistic evidence in favour of any one rather than the other. And there seems to be no non-verbal way of deciding either. People do often utter expressions without knowing the exact meaning of them. This is nothing strange. As Crane says, the expressions of beliefs in a language are always mediated by beliefs regarding, which would be the right words to be used in a situation. Burge, in his linguistic thought experiments, fails to appreciate this. In order to express one’s belief in language one has to have some belief about the use of language. And this becomes a very important point while we make belief ascriptions, because such ascriptions are usually made relying on verbal clues. It is true that there are certain concepts which one could not possess unless she possessed knowledge of public language. But in such cases, Crane claims, ‘the relations between a community’s linguistic practice and an individual’s behaviour have to be mediated by their intrinsic properties’16. And if Burge accepts this, then it would go against his externalist commitments as shown in the general arguments against the twin earth thought experiment by Crane.
Having considered the more specific problems, we might now go over to the more general responses to the externalism from the internalists. Perhaps the most serious reservation we find among philosophers against externalists is that a non-internalistic understanding of mental states is of no use to a scientific psychological theory. Burge focuses on arguments which intend to show that non-internalistic mentalistic ascriptions cannot play a systematic and significant role in psychological explanations. One might think that there may remain some difference between theoretical discourse in scientific psychology and the mentalistic discourse in folk or common-sense psychology. And such differences survive even when ordinary discourse is refined to cater the demands of a scientific theory of mind devoid of vagueness. But it still remains a question as to whether we have to let go of our folk or common-sense intuitions and opt for a purely scientific psychology which, according to anti-externalists, have got to be internalistic.

Tyler Burge is not saying that one has to assume that psychology ought to maintain touch with common sense discourse. What he says is slightly stronger, viz. it will as a matter of fact maintain a contact, and that is how psychology is; and, not only that, but that is how (scientific) psychology ought to be. What I mean by this is that not only does Burge accept that psychology is externalistic, but also that he thinks it ought be externalistic. He is saying that as psychology is committed to the representational nature of intentionality, it is actually about how the mind interacts with the world. In taking psychology as it is rather than

---

how it ought to be, Burge is attempting at an explanation of mental acts. The facts that are commonly accepted by psychology, like people have memory, beliefs, desires, have the ability to categorise things, draw inferences and chose to act and act in distinctive ways by being motivated by desires and beliefs, are the facts which are attempted to be explained in psychology. Psychology attempts, as he says, 'to find deep regularities in these activities, to specify mechanisms that underlie them, and to provide systematic accounts of how these activities relate to one another' \(^{17}\). Burge further claims that in so explaining or attempting to explain, psychology has to make use of 'intentional content'. And such intentional contents can be individuated only in relation to the entities in the world with which they interact. However this individuation of the content is finer grained than the individuation of the entities outside. Burge then goes on to say that he assumes internalism to be \textit{prima facie} wrong about psychology, as it would surely fall prey to the Putnam-kind of thought experiments. And consequently psychology and the language of psychology cannot be purely internalistic. Internalism, says Burge, has got to be revisionary about the language of psychological theory without having any good reason for being so revisionary.

In order to counter Burge and externalism in general, it seems that internalism has got to do two things: 1. The language of psychology has to be revised as the presupposition of science is such that it should be or is internalistic; 2. An explanation of new internalistic language has to be given so as to capture the genuine commitments of the science. Burge thinks that none of these can be accomplished by the internalists.

\(^{17}\) In Burge's 'Individualism and Psychology', p. 4.
We can now move to the more general criticisms against the externalist thesis on the assumption that the twin earth thought experiments do indeed pose a threat to the internalist thesis. We find internalists coming up with a serious argument against Burge in which they turn the thought experiments just against the externalist thesis, which those very thought experiments were designed to prove. The internalists would say that the externalists are missing a very significant point when they say that the two protagonists in the thought experiments do not really have anything psychologically in common. What they are missing is the significance of the fact that the world appears to the twins exactly the same. If they were to switch places, they would not notice it at all. And so we have found philosophers like Segal saying that folk intuition does not unequivocally support anti-internalism. So the argument that Segal is putting forward is that mental states could be individuated by their narrow, individualistic content, since mental states are individuated by their role in psychological explanation and such narrow individuation would be enough for the behavioural explanations that we make in folk psychology. The example that might be used to demonstrate this is as follows: Suppose Ed and his atom-to-atom identical twin in the twin earth are both saying ‘A bear is about to attack me’. However, in the case of Ed, unfortunately, there is a bear, whereas in the case of his twin, call him Ted, there isn’t a bear in the vicinity. According to Segal, we can well individuate the content of their beliefs in terms of what goes inside their head and which has nothing to do with what their environment is like, in the relevant respect. We may even succeed in making all the

---

behavioural explanations about Ed and Ted, i.e. we can predict that both would roll up into a ball, and they both do roll up in a ball.

What we ought to attend to in Segal's argument is: What kind of an argument are we giving when we are individuating the content of Ed and Ted's beliefs and making behavioural predictions about them? We are doing something like this:

If Ed and Ted both say 'A bear is about to attack me', they both will roll up into a ball. As those who say that they are about to be attacked by a bear usually believe that they are about to be attacked by a bear, and they also believe that the only way of saving them from being so attacked is to roll up into a ball. And, if they do not desire to be attacked by a bear they would roll up into a ball.

But what makes us so sure that those who say 'A bear is about to attack me' believe that a bear is about to attack them, and not anything else like a rabbit is scampering by? What makes us so sure is the success of the method we follow while we radically interpret other people's speech and ascribe beliefs to them. It is perhaps true that in this very case, we really don't need to consider the environmental/external circumstances that might have prompted Ed and Ted to make such an utterance, in order to interpret them. But in making radical interpretation and ascription of beliefs, one cannot in the long run ignore the external circumstance in which the subject is situated and the utterance is made. When we first start interpreting other people's speech and ascribing beliefs to them, we do something quite similar to what a field linguist does when she visits a tribe whose language she wants to interpret. We,
like her, have to look at the external circumstances which are invariably linked with the utterance in question and try to construe the meaning out of that. If we consider all this, then ascription of beliefs and interpretation of speech has to be, it seems, in the first instant externalistic. This is so because in the first instance the interpreter has to look into the external circumstances of the utterance to decide what were the beliefs that the speaker had when she was making those utterances. Consequently, one might say, in making behavioural explanations the kind of argument we usually use are such that in them mental states have to be individuated, at least in the first instance, widely.

At this internalists like Segal might say that in the case mentioned above we are confusing the role of the external circumstances in the argument we use in behavioural predictions. The external circumstances are the evidence which we need to use in individuating content. But they are not the individuating properties of the mental states, they do not constitute the identity conditions of the mental state. I am not quite happy with this answer and shall take it up again while I discuss what Crane says against Burge's supposition that Individualists are conflating individuation with causation.

Burge starts by articulating the general argument against externalism and the thought experiment put forward in its defence. The main thrust of the argument is that the two individuals in our thought experiment, i.e. the main protagonists of our story, or the protagonist in the actual and the counter-factual world, are functionally and physiologically identical and hence their behaviours too are identical. And so, psychology (which is the science of behaviour) has got to give a
uniform explanation of the two cases. This argument surely assumes that the behaviours of the protagonist(s) are indeed identical. Burge says that this assumption cannot be made. It can be regarded as true only under a very restricted conception of behaviour, viz. that it is identical with bodily movement. But the notion of behaviour which we find in psychology is richer than this over-restricted notion. This restricted notion is not rich enough to meet the demand of a psychological theory, and if we still go on to adopt this restrictive notion, then we would be restricting the domain of psychology itself. And once we take a richer notion of behaviour, it no longer seems obvious that the behaviour of the two protagonists would be identical. This is particularly so because such a notion includes instances of behaviour which are intentional acts. In psychology we draw evidence from people's verbal and non-verbal behaviour, much of the specifications of which, says Burge, simply have to be intentional, non-internalistic and many externalist accounts of behaviour are relational. And the thought experiments show exactly this. If we take notice of what psychology takes as behavioural evidence, then we would at once see that the individualist's argument is surely ignoring obvious aspects of psychological practice.

Burge seems to assume here that internalists have to take a restricted view about behaviour. This however does not follow from the internalist's rejection of the externalist thesis that there are mental states which have to be individuated widely. Internalists can and do talk about narrow intentional behaviour, and this need not be identical with bodily movements.

---

19 A point Segal made in his Content and Externalism lectures referred before.
20 We shall be discussing this in detail in the third chapter.
Burge finds it wrong to characterise psychology merely as a science of behaviour. We surely cannot characterise a science by the kind of evidence it takes into consideration, rather we characterise it by its subject matter. For instance, if we want to make a correct characterisation of a theory of vision then would it suffice to take it just as a science of behaviour? This wouldn't do, because a theory of vision has got to explain how people succeed in seeing objects in their environment. And in explaining this what we are trying to do is to give an explanation of the relation between a subject and a physical world.

There is however a slightly different argument that internalists offer which needs to be considered. The argument is as follows: that which determines our behaviour is something which supervenes on the states of the brain. And so propositional attitudes which we regard as determinants of behaviour must supervene on states of the brain. Burge is of the opinion that this argument can be inverted. What he wants to show is that not all the causes of our behaviour supervene on the brain. He is giving us two metaphysical reasons and one epistemic or methodological reason for saying that the individualist's argument cannot be true.

Those who make the claim of supervenience often take it to be a trivial consequence of materialism. But we should be clear about one thing; in order to answer the question of what supervenes on what, it is not enough to know what such entities (among which the relation of supervenience holds) are made of. We also need to know how the relevant entities are individuated. So, if we have a mental event, say \( m \),

---

21 This too will be discussed in the third chapter.
which is individuated partly by reference to normal conditions outside a person’s body, then, no matter whether this \( m \) has material composition, \( m \) can well be varying even though the body remains the same. The only way of refuting this would be to show that mental events are individuated in a way such that they supervene on the brain events.

The second metaphysical point that Burge is making is that the internalists conflate causation with individuation. Internalists are of the opinion that nothing can ever causally affect behaviour unless it affects causally the local states of the subject’s body\(^{23}\). And they criticise the non-internalistic thought experiments to be ignoring this basic insight, because the thought experiments talk of environment differently affecting a person’s mental state and behaviour without affecting their body differently. Burge says that the internalists are conflating causation with individuation. Variation in the environment may not causally affect the subject’s body, but can surely determine the individuation of the intentional processes that go within the subject and the way the subject is acting. For example, let us suppose that nothing can causally affect breathing unless it causally affects the local states of the lungs. But it does not follow from this supposition that in order to individuate the lungs and the sub-events of respiration, we have to individuate them only as supervenient upon the chemically described objects and events that compose them. In the same way, we may accept that a person’s mental states and behaviour can be affected causally only through local causal effects on their bodies, but still claim that mental events may be differently individuated though they have the same chemistry.

\(^{23}\) In his ‘Individualism and Psychology’, p. 16.
In his 'All the Difference in the World,' Crane says that Burge's criticism to the effect that the supervenience claim of individualists confuse individuation with causation is unsound. He says that in associating causation with individuation the internalists are not confusing one with the other. What they are doing is that they are individuating properties by their causal powers. Crane feels, like many others, that in order to have a science of the mind we need to individuate mental states in terms of their causal powers. We shall discuss this view in detail in the third chapter when we consider Fodor's criticism of externalism.

Burge now goes on to give an epistemic reason for regarding the internalist argument as wrong. He says that the internalist's thesis that none of the propositional attitudes which are regarded as determinants of one's behaviour can differ unless some of the brain states of the person differed, is a modal generalisation which is grounded neither in judgements about particular cases, nor in the interpretation of actual explanatory or descriptive practices of psychology. The point that Burge is trying to make is that in order to know about individuation or in order to find correct individuative methods, we have to fall back upon the nature of explanation and description in psychological discourse. Internalism, Burge thinks, does not establish, but rather presupposes that explanation and description in psychology has to be internalistic. One might, however, argue in favour of internalistic description and explanation by saying that according to higher level theories of natural science, entities that are postulated by psychology have to supervene on entities postulated by physiology. Or one might try to demonstrate the inadequacy of non-internalist psychology by showing them to be very similar to unsuccessful theories of the past. Burge says that the first
attempt to rule out non-internalist explanation and description is plagued by two difficulties: firstly, natural science is so different from human science that transferring of principles from natural science to human science does not seem to be warranted. Moreover, there may well be instances of natural sciences which are not obviously internalistic. Burge then rules out the second way of countering the claim of non-internalistic explanation and description in psychology on the ground that the previous unsuccessful theories which are claimed to be similar to non-individualistic theories are not as a matter of fact similar to them.

The last argument against externalist thought experiments that Burge considers is that the thought experiments talk of counterfactual situations which are not empirically plausible. But this kind of an argument can hardly be taken seriously. The whole thrust of arguments with considerations of many such counterfactual situations does not depend on taking them as empirically plausible. One is taking into consideration what is logically and theoretically possible, and attempting to make room for them in a general theory.

As we will discuss in the fourth chapter, the main argument against the Burgian thesis is that it cannot accommodate both that causation works through intrinsic properties and that intentional states work causally. We could probably extract similar lines of thought from Fodor when he says that the formal properties of representation

24 These claims however don’t seem to me to be too obvious, they need to be defended. But I am not going into that.
supervene on the intrinsic properties of mental states. He says that it is true that one could not affect the causal powers of a person's mental state unless they affected her physiology, i.e. affecting thus through intrinsic properties. And the world outside is no intrinsic part of the subject. Burge's response to Fodor is somewhat frustrating; he says that Fodor is begging the question by simply assuming the truth of supervenience. But as Crane points out, the importance of Fodor's objection is that whatever the causally mediating properties are, they have to be intrinsic. And to counter this we would have to, in some way, show that the external object is in some sense intrinsic to the thought or reject that whatever the causally mediating properties are they have to be intrinsic. But then what exactly would count as an intrinsic property?

At this point I would like to explicate the real significance of the internalist's emphasis on the significance of the fact that the world appears to the two protagonists in the thought experiments in exactly the same way. We have to keep this in mind when we consider the internalist's response to the kind of argument that Burge has brought against the internalist's conception of narrow content. The internalists might have a strong point to make here. The internalists might say that in the individuation of mental states we should not rely upon the 'narrow' content which Burge shows depends upon some relational properties of the state. We should depend solely upon how the mental states feel like to the subject in question. In other words, we have to see what it is for the subject to be in the mental state in question. And if we look into this, then we have to rely on no relational properties of the mental states at all. All we need to see is how it feels for the subject to be in such a mental state. And this property is purely internal and it is also what changes with the change in the physical properties of the
look into this, then we have to rely on no relational properties of the mental states at all. All we need to see is how it feels for the subject to be in such a mental state. And this property is purely internal and it is also what changes with the change in the physical properties of the individual and it is what is common to the twins in the thought experiments. On this view, we have to depend solely upon what it is like for the subject to be in a certain mental state. We shall discuss this view in detail in the next chapter. This view seems to bring out a possible assumption of the internalists, viz. that the subject can never be mistaken about her own mental states.

1.4 Marr's theory of vision and the externalist-internalist debate

We may now go over to David Marr's theory of vision and its role in the externalist-internalist debate. Tyler Burge in 'Individualism and Psychology' tries to show that not only in philosophy does anti-internalistic intuitions work, but they work also in other scientific psychological theories. And he takes David Marr's theory of vision to be an instance. What Burge claims when he says that Marr's theory is anti-internalistic is that there is some deep, necessary 'individuative relation between some of the representational contents attributed by the theory of the visual system and aspects of the extra-cranial environment'. We however find Gabriel Segal trying to show that Marr's theory need not be interpreted this way, in fact he claims that the straightforward and reasonable reading of Marr's theory is perfectly internalistic. And consequently Marr's theory of vision would count as an evidence in favour of internalism. I shall first try to state what Burge and Segal say about Marr's theory and its role in the externalism-internalism debate.
and then re-iterate in the light of this study that the real issue at hand has
got to be something else\textsuperscript{28}.

Marr's programme is to understand how descriptions of the
world can be efficiently and reliably obtained from retinal images. The
questions that he tried to answer are: What kind of information does the
human visual system represent, what kind of computations does it
perform to obtain this information, and why? How does it represent this
information and how are the computations performed and with what
algorithms? How are these specific representations implemented in
neural machinery? Consequently Marr's theory has three parts. The first
part is known as the computational theory. In this part the visual process
is characterised as a mapping from one kind of information to another
kind of information. The abstract properties of the mapping are defined
in a way so that it can be demonstrated that the mapping is appropriate
and adequate for the task at hand. The second part is the theory of
representation and algorithm. In order to make the information explicit,
we need to have a representational system in which the information can
be represented and the computation must be performed by some
algorithms. And the third part is the theory of hardware implementation
where how the system is physically realised is described.

The computational theory of vision treats our visual system as a
machine which yields three dimensional description of the world once
given pairs of retinal images as inputs. A computational description of
the retinas are given, viz. 'grey arrays' and they represent the light
intensity values in a two-dimensional co-ordinate system. Now what is

\textsuperscript{28}See Segal's 'Seeing what is not there', Philosophical Review, 98, 1989, p. 189.
the computational problem? It is this: given grey arrays as input, how does the system construct the complex description of 3D objects that are its outputs?

To start with the system should describe the patterns of light and shade on the retinas and that is given in terms of representations which can be obtained about the structure and changes in the images. From these representations we can get certain features which are useful in constructing higher level representations. After this certain computations are made to construct a 'primal sketch' of the surface outside the viewer. On the basis of this primal sketch the system then goes on to construct a 2.5D sketch of the viewed surface and from this finally the construction of the 3D model representation. And Marr says that process of one stage of representation to the other is made by means of inference. And so it is essential to find out what proper assumptions are brought to bear by the system, that allow it to make the inferences. The assumptions built into the system is very important, as the visual system is supposed to provide an accurate representation of the world outside. Another important thing to be considered is that as the visual system infers from the structures of grey arrays the representations of 3D objects in objective space, the nature of lower-level descriptions should be such that they are recoverable from levels below and can be used in levels above. And so the process of constructing a representation is regarded to be inferential and so the theory of vision is a theory of inferences. And each of these inferences needs to have an input from the level below, an innate assumption and an algorithm by means of which an output

---

28 In discussing Marr, I will be following Segal's exposition of Marr's theory, as we find it in 'Seeing what is not there'.

representation is computed. And so each attribution of representation requires a bottom-up account.  

Again since we are looking for a bottom-up account, the amount of computational activity required would vary as the specificity of the informational content required varies. And the 'information that is made explicit at each stage below the top-most is carefully chosen to suit subsequent computations. This would be the same even in the case of the highest representations, i.e. the 3D model representations as they too are again used for the recognition of shapes. In fact the role that a particular representation has in the theory is articulated by its place in the sequence of inferences that begin with the grey arrays and culminate with the recognition of shapes of objects.

The other most important point is that the attribution of representations has to be checked against empirical evidence. For example important evidence may be provided by the precise conditions that cause illusions.

As this is what Marr's theory of vision says, Burge claims that it has got to be anti-internalistic. The argument that Burge gives in favour of his view is as follows:

1. The theory is intentional
2. The intentional primitives of the theory and the information they carry are individuated by reference to contingently existing physical items or conditions by which they are normally caused and to which they normally apply.

---

29 Segal, G. 'Seeing what is not there', p.194
3. So if these physical conditions, and possibly attendant physical laws, were regularly different, the information conveyed to the subject and the intentional content of her visual representations would be different.

4. It is not incoherent to conceive of relevantly different physical conditions and perhaps relevantly different (say, optical) laws regularly causing the same non-intentionally, internalistically individuated physical regularities in the subject's eye and nervous system.

5. In such case, (by 3) the individual's visual representations would carry different information and have different representational content, though the person’s whole non-intentional physical history might remain the same.

6. Assuming that some perceptual states are identical in the theory in terms of their informational or intentional content, it follows that internalism is not true for the theory of vision.

Burge seems to claim in 2 and 3 that on Marr's theory of vision the intentional content of visual experience is individuated by reference to external environment. And 4 and 5 resonates of twin earth intuitions. What is said here is that the way that Marr says that the representational content is individuated is very different from what any individualist would say it was. In order to explicate Burge's position, Segal gives yet another twin earth thought experiment. Let us take Visua an earth-inhabitant and Visua* a twin earth inhabitant both having exactly the same individualistic histories, i.e. they have gone exactly the same sequence of grey arrays and their brains have undergone the same changes seen internalistically. But as usual their environments vary in a way that does not make their narrow content change. Burge would claim
that Marr uses the contingently existing items of the environment to individuate the representational content of Visua and Visua*. The items that cause Visua*’s grey arrays are systematically different from the ones that cause Visua’s, though the grey arrays so caused are exactly the same.

Segal tries to challenge this very interpretation of the thought experiment and tries to defend the view that Marr’s theory is not anti-internalistic. Segal says that Burge’s claim is acceptable only under a restrictive interpretation of Marr’s theory of vision and he offers a liberal interpretation which would demonstrate that Marr’s theory is not anti-internalistic. To start with, Segal points out that the difference in the representational content of Visua and Visua* is partly due to the innate assumptions of the system and so though they may be subject to the same grey arrays yet they might differ in their innate assumptions. Segal says30 that Burge seems to assume that Marr is subscribing to the view that what a particular representative type represents is normally what causes it. And this surely is the counterpart of Putnam’s Causal theory of reference in the theory of vision. That is the thesis that the content of representation is determined partly by the underlying nature of the cause. Consequently Visua and Visua* will have representational types each of which would represent its typical cause, but as they have different causes of their representations their representational contents are bound to differ. As Segal has pointed out, their innate assumptions too should differ. Suppose the constraints on earth are C_e and the constraints on twin earth are C_t. And the constraints are such that the same grey array, given C_e, would make us infer object O_e and, given C_t,}

30 ‘Seeing what is not there’, p. 201.
would make us infer object $O_t$. On Burge's view, Visua* correctly sees in twin earth as on twin earth the grey arrays $G$ are typically caused by objects $O_t$. So correctly seeing in twin earth would require the representation of the form $R_t$ and not $R_e$ as is required on earth. Segal does not think that this is acceptable. This is so because he contests the thesis that from the mere fact that grey arrays $G$ are caused by $O_t$ that the correct representation constructed from $G$ must be representations of $O_t$ and should be of the intentional type $R_t$. And Segal then says that in a particular case of a representation that is 'satisfied' by both $O_e$ and $O_t$ will be correct in twin earth. And if the representations constructed by Visua and Visua* are of the neutral type $R_n$, which would be satisfied by both $O_e$ and $O_t$ then they would see correctly even though they have the same representations. This is counterpart of what Burge calls the thalidomide concept which is shared by the protagonist in Burge's linguistic thought experiment. And it can well be the case that the systems shared the common neural constraint $C_n$.

Segal says that Burge's restrictive reading inspired by causal theories of reference may be true of certain parts of folk psychology, but there is no reason to suppose that it would be true of the theory of vision. The real issue is whether or not under the restrictive interpretation, the theory of vision can provide the best possible explanation of vision.

Segal then goes on to show that an individualistic liberal interpretation of Marr's theory can provide us a satisfactory explanation of vision. In doing so he elaborates Visua and Visua*'s story a little further. Suppose Visua* is brought to earth. And Both Visua and Visua* are experimented on in a laboratory. First, circumstances are created so
that \( C_e \) is violated and \( C_t \) is obtained. Then we subject the twins to the object type \( O_t \) to cause grey arrays \( G \). In a second experiment we arrange the circumstances in a way that constraints \( C_e \) obtain and \( G \) is caused by \( O_e \). Under the restrictive interpretation of the theory, in the first case Visua* correctly sees \( O_t \) whereas Visua has an illusion and in the second case Visua* has an illusion of \( O_e \), whereas Visua sees correctly. Under the liberal interpretation they both see all the ‘distal causes of \( O_n \)’. As we have already seen, while discussing Marr’s theory, the only motivation for opting for attribution of \( O_t \) and \( O_e \) rather than a neural one would be that one rather than the other is required to find some higher computational process into which this content would be fed. And there doesn’t seem to be any such motivation in favour of the restrictive interpretation. As Segal says: ‘The general message is that we can never suppose that a mere difference in distal cause, however regular, however ubiquitous in evolutionary background, must make a difference to representational content. To demonstrate that the differences between the twins’ environments showed up in perception, we would have to demonstrate that their representations were specific enough to distinguish the two sorts of distal cause. Given the identity of discriminative abilities, this could never be shown.’\(^{31} \)

### 1.5 A puzzling situation

At this juncture I would like to make a few general observations about the debate between externalism and internalism. It seems to be very hard for me to see any common ground on which the externalist-
internalist debate can be decided. It is not the case that each and every argument that may be brought against Burge’s thesis can be regarded as question begging. No thesis, if it is to be a serious one, can be such that it almost logically rules out the possibility of there being real objections to it. When we see internalists and externalists arguing over what should be regarded as the individuating property of mental states, then it can seem that the views that each of them are arriving at is rather arbitrary. Views which are justified (if at all) by either internalistic intuitions or by externalist intuition are such that one who lacks any of those intuitions fails to see why one rather than the other view is acceptable. Perhaps the only apparent common ground between internalism and externalism is the twin earth intuition. But Segal and Crane (at least) dispute the significance of these intuitions. Crane also says that his rejection of Burge and Putnam’s arguments for externalism in form of thought experiments do not lead him towards the position that intentional states have to be individuated by their narrow content. It is only if one takes the twin earth cases to be posing a genuine problem, that one would opt for a narrow content as a solution to it. He rejects broad content nonetheless. But for different considerations and not because of twin earth intuitions. Perhaps with Crane the real issue between externalism and internalism is one of mental causation, and perhaps it is because of this that he is inclined towards internalism.

Perhaps the real issue at hand is quite different. Perhaps it is one to do with physicalism, or one to do with the possibility of mental causation. It could even be about some issues in the philosophy of language, like the semantics of proper names and natural kind terms,

---

32 In ‘All the Difference the World’
33 We shall see this in the fourth chapter.
whether they involve only reference or reference and object-involving sense, or sense which is not object-involving. May be it is one of these issues rather than anything else which pulls us towards either externalism or towards internalism.

For instance, Segal thinks that external circumstances are just evidence which we need to use in order to discover the individuating properties of mental states, but they themselves on their own do not have such an individuative role. But the reason that externalism seems to be plausible to many is that in a very ordinary sense we do individuate mental states by their wide content. We do recognise or identify beliefs or mental states in general and distinguish them from others by reference to the object towards which it is directed. And this object more often than not is an external object. And hence there is a tendency of individuating mental states in terms of external objects.

But here the internalist may have something important to say. We do have cases of mental states which are about objects which do not exist and these cases pose a serious threat to the externalist thesis. Individuation of these mental states cannot possibly involve their relations with external objects, as there isn't any such external object. However it would be unfair to think that externalists were making the absurd suggestion that we can only think about actually existing objects. Externalists like Burge have said that even in the case of mental states which are about non-existent object, we have to bring in some relation to some external object or other. Burge points this out in his 'Other Bodies'. Here he considers cases of beliefs about water where water does not

\[34\] A point which he made during the discussion on his presentation in the Content and Externalism seminar in Kings College London, 1995.
exist. Suppose someone is having an illusion of water. Burge says that part of what we do when we conceive of such cases is to rely on actual circumstances in which these illusions do not hold — rely on the actual existence of water. We utilise — must utilise, I think — the actual existence of physical stuff and things, or of other speakers or thinkers, in making sense of counterfactual projections in which we think at least some of these surroundings away.35

Perhaps what the internalists want the externalists to appreciate is that the individuation story has to be in some way the same, or invariant for the thoughts which are about existing objects and for thoughts which are about objects which do not exist. The individuation of mental states should be such that it is applicable to both ones that are directed towards existing things and those that are directed towards non-existent things. But the uniform explanation or uniform individuation story which we are looking for should in no way be construed in a way which is in favour of internalism or externalism. Perhaps here again it is an issue at the level of what kind of semantics we are opting for, for empty terms. And we have got to settle that issue first and then the rest would just follow.

It seems that we are facing a deadlock. Perhaps we face it because we are looking for a uniform way of individuation of different kinds of mental states. Both internalists and externalists are doing this. Internalists want to individuate mental states in terms of their narrow contents, as the narrow content is what is invariant to all kinds of mental states. The externalists (like Burge), on the other hand, are trying to demonstrate that even the individuation of the invariant narrow content

35 'Other Bodies', p. 114.
of mental states involves some relational elements. There seems to be an urge amongst both the contending parties to bring all mental states under one single mode of individuation. But perhaps we should have different stories to tell for mental states about existent objects and for mental states about non-existent objects. Or alternatively one could try to construct the individuation story in a way which would be neutral to the kind of object that it is about (external/existing or otherwise). It seems that we can describe thoughts in an externalist way or in an internalist way; the question is: what is the point of these different descriptions?

I feel that the externalist/internalist debate seems to present itself with a kind of deadlock. What I mean to say is that it could be that the externalist and the internalist are opting for their respective views about the mind and the mental, because they have specific views about quite a few things. For instance, views regarding individuation in general, individuation of mental states in particular, views regarding psychology and what is explained in psychology and what psychological explanation ought to be like, views about how mental causation is possible. And it is because they differ in their views about these issues so radically that we find the debate in a state of utter deadlock. I feel that all this happens because the debate seems to be bottoming out into intuitions too early on. And I feel one needs to find out what exactly those intuitions are and why are they influencing the externalists' and the internalists' view so early on. Otherwise, it would seem as though there is no real debate going on amongst the contending parties.

This is why I think that the best strategy to understand the contending positions, and hence the debate, would be to look for the possible assumptions that they might have. In this thesis I want to come
up with a negative defence of externalism by looking into two possible assumptions of the internalists which I shall try to show are not quite acceptable. From the way the internalists have responded to the externalistic arguments based on the twin-earth intuitions seem to show that they have two such assumptions. In what follows I will be discussing why I think that the internalists have such assumptions and also why I think that these assumptions cannot be defended. The two assumptions are:

(a) We can consider two mental states to be the same (i.e. type-identical), if the subject cannot discriminate them.

(b) We can consider two mental states to be the same if they give rise to the same behaviour, ceteris paribus.
Chapter Two
The Identity(-Individuation) Crisis
2.1 Introduction

From the first chapter we have found that it is really difficult to make sense of the externalist and internalist debate. May be, it is because of this, it is hard to see how the debate could be resolved or solved. I think that the way in which the internalist arguments against the externalists run, indicates that they are working with two assumptions/presuppositions and I want to come up with a defence of externalism by questioning those two assumptions. The point that I want to make here is that there are good reasons for thinking that the internalists make these assumptions, and if they do, then they are bound to face some serious difficulties. And if we can find good reasons for questioning these assumptions, then we can build up a strong case for externalism.

The internalist assumptions that I am talking about can be extracted from two major arguments that they bring against the externalists. They say that the external object towards which the mental states are directed, or the external environment in which the mental states are experienced by the subject have no essential (individuative) role to play in the individuation of mental states. This is so because:

(a) We can consider two mental states to be the same (i.e. type-identical), if the subject cannot discriminate them.

(b) We can consider two mental states to be the same, if they give rise to the same behaviour, ceteris paribus.
That internalists might be working with (a) as their assumption is clear from the way they respond to the twin earth thought experiment, as we have already seen in the first chapter. The internalists say that since Ed and Ted would not be able to differentiate amongst their own water and twater thoughts, their thoughts of water are of the same type as those about twater. Again, we have seen that they behave in exactly the same way as a result of their water and twater thoughts, in earth and twin earth respectively. And since their water and twater thoughts give rise to the same behaviour, Ed and Ted must be in the same kind of mental states in both the worlds, in spite of the fact that one of the worlds has water and the other has twater.

This line of thought amongst internalists can be found in the writings of Gabriel Segal and Simon Blackburn. This approach has been condemned in the literature as 'Cartesian', since it endorses that 'we know what it would be like, from the inside, to be in the position of my subjects in the various possible cases. And when we think what it would be like, we can see that the thoughts we would have under substitute and empty possibilities would be the same, regardless of how they were caused'. The reason this approach is called Cartesian is that it implies a privileged authority on the part of the subject, which would be in some sense infallible or incorrigible or at least the best possible authority. I shall start my discussion of the presupposition (a), by showing that having subscribed to it (i.e., (a)), internalists can be said to be adhering to Cartesianism, since (a) is a natural consequence of Cartesianism. I shall be discussing the works of Segal and Blackburn as

---

36 See Segal's 'Return of the Individual', , and 'Seeing what is not there'.
38 See McDowell's 'Singular Thought and Inner Space' in Pettit, MacDowell(eds.), *Subject, Thought and Context*, pp. 146-158.
they are subscribing to Cartesianism of the aforementioned kind. In this
connection I shall be discussing the views of Gareth Evans and John
McDowell who have tried to demonstrate that this kind of Cartesianism
is unacceptable.

I also try to give my own reasons for taking (a) to be wrong in this
chapter. The point that I want to make is that the question of sameness
and difference of mental states is not tied to the question whether the
subject can discriminate the states or not alone. Subjective
indiscriminability of mental states is not an infallible mark and
consequently cannot always be regarded as a good evidence for the
sameness of mental states.

This, I shall argue, is because there are general difficulties in
taking subjective indiscriminability to be the mark of identity.
Metaphysically speaking, the idea of identity and the idea of
indiscriminability are distinct. Philosophers have tried to show that at
least in the cases of mental states these two notions coincide. Internalists
can take two stands on this issue. They can opt for a radical infallibility
thesis, which says that as the subject is never mistaken about her own
mental states; if two mental states seem to be same to the subject, then
they are indeed so. The less radical stand is that subjective
indiscriminability is good evidence for the sameness of mental states.
The internalists say that if, what is presented to the subject or what
appears to the subject is the same, then it is a good evidence for the
identity of two experiences or two mental states. What I am trying to do
is to give a rational re-construction of the internalist story in order to
explore what could have been the motivations behind the internalist

39 Blackburn, Spreading the Word, pp. 323-324.
thesis. And it seems to me that (a) can well be one of the general principles that the internalists can be assuming in order to defend their position (as a natural consequence of their Cartesianism). I want to maintain that there is something wrong with this view:

1. If the internalists were to assume the first approach (i.e. the infallibility thesis), they would run the risk of begging the question. This is so because by already presupposing the general principle (that we can consider two mental states to be the same, i.e. type-identical, if the subject cannot discriminate them), the internalists are taking up a particular stand about how mental states are to be individuated; and,

2. It is not the case that, as far as mental states are concerned, the subject’s failure to discriminate the two mental states is a sufficient condition for their being identical, for there are counter instances. The reasons for which it is believed, by some philosophers, that subjective indiscriminability in the case of mental states is a sure mark of identity are faulty. For one thing, these philosophers believe in the infallibility thesis which says that the subject is never mistaken about her own mental states. And even if the internalists do not believe in the extreme infallibilist thesis, they do indeed believe in the incorrigibility thesis. Both these theses heavily depend upon the idea of phenomenal character — an idea which is not obviously acceptable. At least we can say that the idea of phenomenal character is shrouded with mysteries which are not easily solved.

3. Considering all this it seems to me that subjective indiscriminability is not always a good evidence for the sameness of mental states. There are clear instances of mental states where indiscriminability is not a good evidence for sameness of such
states. This being the case it is difficult for the internalists to show that the other mental states are not like those clear instances.

2.2. What it all hinges upon: Cartesianism and Internalism

To make my line of argument run, I shall have to show that the internalists do have (a) as an assumption. And if they are indeed subscribing to such an assumption, then they would be facing a lot of problems. It probably isn’t obvious why internalist might be assuming it, at least not directly from the above characterisation of the internalist thesis itself. I shall now try to show why I think they indeed have this assumption.

I think that if we look at the reasons that could have led the internalists to opt for narrow as opposed to wide individuation of mental states, we will understand that they definitely have (a) as their assumption. In this connection I would like to mention one of the common beliefs that are associated with our understanding of the individuation of any category. We commonly believe that in order to find out the individuative characteristics of any category (in our case, mental states), we have to look for that characteristic which is invariant to all entities which belong to that category (in this case, we have to find the characteristic which is invariant to all mental states). So what is it that is common to mental states in general? We have to be careful when we speak of mental states in general, as conscious and unconscious mental states should be differentiated from one another and what would be true of one might not be true of the other. I am, however, concentrating on conscious mental states. This is so because we are
talking about how one person can distinguish amongst her own mental states. And so of course in this case we are speaking about mental states which the subject is conscious of. There is a motivation for finding what may be regarded as the general character of mental states. And it is this motivation which I think has led the internalists to think that it is the 'how it feels to the subject' property which can be regarded as the individuative character of mental states, since it is this property which can be said to be invariant amongst mental states in general. For instance, this is what is invariantly present in all kinds of conscious mental states, mental states (putatively) with narrow content and mental states (putatively) with wide content alike, mental states about existing objects and non-existing objects alike. And considering this character, the mental states of the twins across the worlds do indeed seem to them to be the same. So Ted has a thought in twin-earth which has the 'how it feels' property which is absolutely the same as the 'how it feels' property associated with the thoughts he has when he is transported to earth.

There is another common belief which has gone into the internalist assumption: It usually seems to be a good reason to take two things as being of the same kind, if we cannot tell the difference. In most cases this really works as a good evidence, unless we are proven wrong by some extra evidence that we didn't previously have. These are perhaps the methodological insights that have made the internalist believe in (a).

Finally there is the Cartesian insight which has put together all these commonly held beliefs to form the internalist assumption, viz.: usually a subject is in the best position to know what her thoughts are like. Let us be a little more systematic and let us now enumerate the possible beliefs that I think have gone into the internalist assumption that we can consider two mental states to be the same (i.e. type-identical), if the subject cannot discriminate them:
1. To individuate mental states we have to look for the common element invariant to all mental states.

2. It is the 'how it feels property' of mental states (i.e. the phenomenal character of mental states) which is common to all mental states.

3. As far as this 'how it feels' property is concerned, Ed and Ted are similar. Thoughts in an environment with water feels like just the thoughts one has in an environment without water. And so Ted cannot tell the difference of his 'water' thoughts in twin earth from his water thoughts in earth.

4. If two things seem to us the same, then it is more likely than not that they are the same.

5. There is a reason other than 4 to support the claim that if Ted cannot tell the difference between his 'water' thoughts across the worlds, then he is in same kinds of thoughts. And it is: We are usually right about our own mental states.

In what follows, we shall see that internalists indeed have these beliefs. We shall be questioning each of these beliefs and we shall thereby be showing that the internalist assumption has got to be wrong.

One reason why it is plausible to suppose that the internalists might be assuming this is as follows: Let us take the twin earth thought experiment. The argument that the internalists bring against the externalists is that since the world appears to the two protagonists in the thought experiments to be exactly the same, there seems to be good reason for thinking that they are in same states of mind. And this is all that matters. It is true that, more often than not, if we cannot tell one
thing from another, then they are in fact the same. Similarly, in the twin earth case, as the subject cannot tell the difference between his mental state in twin earth and his mental state in earth, the subject has good reasons to suppose that he is in the same mental state. And if the twins were to switch places, unknown to them, they would not notice it at all. And since the protagonists cannot discriminate their mental states in one world from the other, the worlds and the objects in it form no essential individuative part of their mental states. They are in the same mental state and if one of the mental states fails to entail the existence of an object in the world or entail anything in the environment in which the mental state is entertained, then so will the other. Consequently, the internalists conclude that externalists are wrong to suppose that mental states are to be individuated in a way so that the very existence of the state entails the existence of the object towards which the states are directed, or make entailments about how the environment is.

This being a general sort of reason for thinking that the internalists indeed adhere to (a), there is another reason which they might have for adhering to (a), which is a result of their Cartesian bent of mind. We can have textual evidence which shows that the internalists are indeed Cartesian in their approach. Being Cartesian they believe that a subject has a privileged access to her mental state is a view that is largely accepted. And so in the twin earth case as the two protagonists are in the best position to judge the character of their mental states, if they cannot tell the difference between their mental states in the two worlds, they must be in the same mental state. And what is it that is same in the thoughts of the twins in the two worlds? The objects that their thoughts are directed to are not what remains the same across the
worlds. And hence their thoughts, widely individuated, are not the same across the worlds. But they are the same, if they are individuated narrowly.

We will now discuss, following Evans and McDowell, that the way that internalists have tried to counter the possibility of there being thoughts (singular thoughts) which are tied to their external objects for their identity, demonstrates that the internalists adhere to a kind of Cartesianism. This Cartesianism is found in the writings of Simon Blackburn and Gabriel Segal. We find both of them rejecting the kind of (Strong) Singular Thought Theory that Evans and McDowell have proposed to defend.

In the article ‘The Return of the Individual’ Gabriel Segal is trying to refute what he calls the Strong Singular Thought Theory (SSTT from now on). This is a theory which was endorsed by John McDowell in ‘De Re Senses’ (DRS from now on)\(^{40}\), ‘The Sense and Reference of Proper Names’ (SRPN from now on)\(^{41}\), ‘Singular Thoughts and the Extent of Inner Space’ (STEIS from now on)\(^{42}\). In doing this Segal has actually come up with a defence of internalism/individualism. The same kind of attempt at refuting the SSTT was made by Simon Blackburn in his *Spreading the Word*\(^{43}\) and he too in doing so has come up with a defence of his internalism. We shall discuss this as well.


Before going into Segal’s response to SSTT\(^4\), let us first try to understand what Segal thinks the SSTT is saying. SSTT is saying the following things:

1. Some singular thoughts are object dependent in the sense that they depend on the physical object in the external world (towards which they are directed) at least partly for their existence and identity. (This definitely is a typically externalist way of viewing certain mental states.)

2. So no part of the content of such thoughts is such that the existence and nature of it is independent of the objects they are about.

3. Two such singular thoughts, if they have different objects, then by virtue of so having different objects, they would also have different contents. (So the sameness and difference of singular thoughts, under this view, depend upon the sameness and difference of the objects towards which the thoughts are directed. Hence here we have a species of mental states whose identity is tied to the identity of their objects.)

4. SSTT rejects the possibility of reference-independent singular thought, i.e., empty singular thoughts. A SST theorist rejects this because SSTT rejects reference-independent content.

Segal is opposing all of this. Segal finds SSTT unacceptable because of the following reasons:

1. SSTT is incapable of fully accounting for the psychological role of singular thoughts.

2. Segal says that to make these thoughts appropriate for psychological explanations, we have to reject that the referent of the thought is in any way relevant to the type-identity of the thought. He says that a ‘taxonomy must collect appropriately similar singular thoughts about different things, and include in

the same type even thoughts the singular component of which lacks a referent altogether.\textsuperscript{45}

3. Segal thinks that SSTT fails to explain actions of subjects of empty thoughts, i.e. thoughts which seem to have some reference but does not actually have them.

4. Finally he thinks that SSTT undermines itself.

5. Segal says that there could be milder versions of SSTT but even those versions are inadequate.

Segal says that in psychology we must individuate thoughts by their contents. But contents, unlike what SSTT says, can well be reference independent. And he argues in favour of this content rather than the one endorsed by the SSTT.

We can find arguments against SSTT in Fodor which we will discuss in detail in the third chapter. We will discuss Segal and Blackburn here as they both seem to me to be taking up the first assumption.

To understand McDowell's SSTT we have to understand Russell's notion of singular thought. McDowell's SSTT is a blend of Russelian and Fregean insights. Russell regarded singular thoughts as object dependent. Russell was extremely strict about what could be regarded as a singular thought. He imposed a restriction on singular thoughts (which is known as Russells' Restriction). We find Russell saying in the chapter on Knowledge by Acquaintance and Knowledge by Description, in his \textit{The Problems of Philosophy}, 'Every proposition which we can understand must be composed wholly of constituents with which we are

\textsuperscript{45} Ibid., p. 40.
acquainted" (we can call this the principle of acquaintance). The principle that Russell is suggesting about singular thought is that it is the only thing with which one can be 'acquainted with' (things about which one can have knowledge by acquaintance as opposed to knowledge by description), which can be objects of one's singular thoughts. If we apply Russell's restriction to singular thoughts, then we will be bound to say that the only genuine object for a genuine singular thought will be 'sense data ... and the thinkers of the thoughts themselves'. This is so because for Russell the paradigm case of acquaintance is perception.

Both McDowell and Gareth Evans have found that this principle is far too strict. Following Evan's suggestion one can make this principle less stringent by reading 'Russell's notion of acquaintance into a simplified form of Evans' account of perceptual demonstrative mode of presentation'.

We can now discuss McDowell's position regarding Russell's Principle and see how he thinks that a strict adherence to it by the internalists would lead to Cartesianism. There seems to be reasons for thinking that Russell's Principle as conceived by Russell himself is unacceptable, and that it is this principle which makes the Russellian scheme distinctively Cartesian. This has been brought out in the discussion of the principle by McDowell. This probably is the key point at issue for us. I feel that the internalist thesis too is distinctively Cartesian. Let us see what McDowell means when he says that the

---

46 See Russell's *The Problems of Philosophy*, Chapter on Knowledge by Acquaintance and Knowledge by Description.
49 'Singular Thought and the Extent of Inner Space', p. 141.
Russellian thesis is Cartesian and let us see what he means by this and why he is opposed to it. I shall also try to say why I find the internalist thesis Cartesian and how we find streaks of this Cartesianism in Segal and Blackburn’s work, especially in their critic of McDowell’s SSTT.

What Russell wanted to express by his principle of acquaintance is that in order to make a judgement about an object, the object should somehow be the one that the person in question has in mind. That is, the person has to know who or what or which the object is. But this perhaps is too much to ask of a person. The knowledge required should not be characterised as ‘knows which’. But maybe it is not quite right to attach this requirement to knowledge by acquaintance. One may quite well be able to make a judgement about an object which she knows in acquaintance without having the faintest idea about which the object is. (As Evans says in his discussion on Russell’s Principle, one may have the same kind of knowledge about oneself without having any clue to who one really is!) But nonetheless, Russell holds on to this principle. The reason for which Russell restricts the objects of singular thoughts to sense data, the subject and its inner states, is that only in these cases the subject can be said to know which these objects are. Now why is this so? This is so because, following a Cartesian and Russellian line, it is believed that one can never be mistaken about the sense data that she receives in experience, and one can never be mistaken about oneself and one’s own inner states. And so if we lift Russell’s restriction we open the possibility that a subject may be in error about the contents of his own mind. Russell of course would never agree to this, and McDowell says that ‘the ultimate basis for Russell’s restriction is a conception of the

---

inner life, and the subject's knowledge of it, which it seems fair to label "Cartesianism". As we find in Evans, in order to understand the significance of Russell's restriction we have to understand this Cartesian streak in him. To understand what Russell's regards as objects of acquaintance, we have to realise that for Russell all 'the items with which a subject may be acquainted have what may be termed the Cartesian property: it is not possible for a subject to think that there is an item of the relevant kind with which he is acquainted (and hence to think that it is possible to essay a thought about it) without there being such an item'. On a Cartesian scheme we accept a sort of transparency thesis about our inner life. It is thought that our inner happenings are transparently available to us and hence we have an incorrigible, infallible, and direct access to it. Russell has been influenced by this Cartesian view and says in *The Problems of Philosophy* that it is always the 'subjective things' which 'are the most certain'. McDowell says that there are reasons for having epistemological worries about such a Cartesian principle. Such a principle in Descartes has made the threat of the loss of the external world greater than ever. It is generally felt that there is a danger in restricting the realm of knowable truths to the inner life of ours. We shall see how we can be equally mistaken about our inner life as we can be of our external environment and we can at times be equally right about the environment as we can be of our inner states.

The Cartesian picture makes the inner realm self-standing, unaffected by the external environmental circumstances. This Cartesian

---

51 'Singular Thought and the Extent of Inner Space', p. 145.
52 Ibid., pp. 145-146.
53 *The Varieties of Reference*, p. 44.
54 P. 18.
55 For a detailed discussion of this see the fourth section of McDowell's 'Singular Thought and the Extent of Inner Space'
picture has become appealing to modern day science. The motivation of making a science of psychology (as we have already mentioned before) has become influential in shaping our theories about the mind and the mental. And as we have already mentioned in the first chapter, and as we shall see in the third and the fourth chapters, it is believed that in order to make a science of psychology, we have to individuate mental states in a way that their 'intrinsic nature can be described independently of the environment'\textsuperscript{56}. The independence and the autonomy of the inner realm has led to the belief that 'how things seem to one is a fact, knowable in a way that is immune to the sources of error attending one's capacity to find out about the world around one'\textsuperscript{57}. The main reservation that McDowell has with Cartesianism is that it broadens the gap between the inner and the outer by giving the inner a special status of being self-standing and immune to sources of error. This will breed a kind of scepticism which is not healthy. McDowell would like to lift this restriction laid by Russell and say that not only are singular thoughts about our own selves and sense data and our own mental states, but they can be about objects outside us. But the internalists would say that if we lift the Russellian restriction and talk about object-dependent content, then such a content would not constitute 'an intrinsic feature of states or episodes in the self-standing inner realm'\textsuperscript{58}. It is this Cartesian view which we find in the writings of the internalists. And it is this view which I want to question here. And in doing so we have to go back to Segal and Blackburn again and see how they defend their Cartesianism as against the anti-Cartesianism of McDowell.

\textsuperscript{56} 'Singular Thought and the Extent of Inner Space', p. 152.

\textsuperscript{57} Ibid., p. 154.
We find internalist writers like Simon Blackburn opposing the McDowellian view about Russell's Principle. Blackburn has tried to show that if we lift Russell's restriction from Russell's Principle, then we would be left with thoughts which might seem to be object-dependent, but they really are not. In his critic of McDowell, he has demonstrated the Cartesian streak in his brand of internalism. And I intend to argue that this Cartesianism is faulty and leads internalists to a faulty assumption (viz.(a)).

Let us see why Blackburn opposes the thought that we can lift Russell's restriction. He tries to demonstrate that it would be wrong to lift this restriction and his way of showing it is by what he calls 'spinning possible worlds'\(^59\). Blackburn, following Russell, tries to argue that the thoughts which we usually take to be singular and hence object-dependent are not so, as they can be reduced to definite descriptions which can well be identity-independent. Blackburn is saying that if we lift the Russellian restriction, we will have to say that in those supposed cases of singular thoughts and the utterances associated with them, the subject could have those thoughts and understand the utterances only if it is known to the subject which object is involved. Blackburn will show, by the help of his 'spinning possible worlds' examples, how this is not required for the having of the thought and the understanding of the utterance associated with it.

He starts by speaking of the Putnamian twin-earth thought experiments and then he gives us his own thought experiments as well. Under the externalist thesis, since in the case of Ed it is water (H\(_2\)O)

\(^{58}\) Ibid., p. 156.
\(^{59}\) Spreading the Word, pp. 310 – 316.
which has caused him to have the water-thought and has prompted him
to make the water-utterance, in the absence of it neither he nor anyone
else can be said to be able to have a water-thought or make a water-
utterance. For Blackburn this is not true. He says that to understand Ed’s
utterance or to get at his thoughts we have to ask ourselves the
question: 'What might have been the same, about the subject, even had
it been a different object causing the very same differences in him?' (As
we have already said that there is a tendency of finding out the common
element in mental states in order to individuate them. Blackburn
demonstrates this belief here.) Blackburn considers this to be a very
important question, since an answer to this would throw light on the
relation between our thoughts and utterances to their objects. Blackburn
says that philosophers are apprehensive of the possibility that the same
experience might be caused by different objects or even by no objects at
all. The uneasiness that philosophers have with this is because this
increases the gap between our experience and the world. This is
something that we have discussed when we spoke of McDowell’s
reservations against Cartesianism. Blackburn thinks that such a fear of
loss of the world could not be eradicated by just denying the possibility
of the same experience being caused by different objects or no objects at
all. The importance of the kind of question that we have just spoken of is
that we need to know what is same and different in the twin cases, and
how these sameness and differences affect the nature of our thoughts
and utterances. What Blackburn intends to defend, like a true internalist,
is that it is what is same amongst the two twins which matters most in

60 Ibid., p. 311
61 This is a kind of question often asked in Philosophy. I feel that the answers to these
kinds of questions are often not as enlightening as they might seem to be. There is a
tendency to bring in all kinds of experiences under one single theory in the literature.
This is also found in the philosophy of perception. We shall be discussing this later.
fixing the identity of their thoughts and which determines the understanding of their utterances. His ‘spinning possible worlds’ examples are designed to demonstrate this. Let us now discuss his thought experiments.

Blackburn asks us to imagine of two musicians who both have heard the most amazing violin playing and they decide that they ought to buy it. Suppose in one possible world ($w_1$) it was Stradivarious1 that they had heard; and in a substitute possible world ($w_2$) it was Stradivarious2 that they heard; and in a third possible world ($w_3$) there is no violin at all that they heard; it was just a computer simulation. Blackburn points out that if we look at what it is like for the subjects, then we will realise that the three situations are the same. They have the same thing running through their head in these three situations. Maybe they think in all these circumstances ‘Ah! This violin is divine!’, and they may be saying in all these circumstances ‘I must have that violin’.

In both these cases the demonstratives that we are using are usually regarded as singular terms, but whereas in $w_1$ they refer to Stradivarious1, in $w_2$ they refer to a different thing Stradivarious2, and in $w_3$ they refer to nothing at all. But how can genuine singular terms be like this? Blackburn emphasises that the subjects’ understanding of their own thoughts and utterances across the possible worlds will remain the same. And hence their understanding of their thoughts and their utterances are not tagged onto the objects and references of their thoughts and utterances. And so Blackburn concludes that ‘we need another ingredient than reference to locate their understanding’

The understanding should be located in the invariant element of the
experiences across the worlds. And this is nothing but the way in which the objects appear to the subject, i.e. the mode of presentation or sense. It is this which determines what the subjects understand across the contexts. And this is universal, shared by all worlds, one in which there is only Stradivarious1, the one in which there is only Stradivarious2, and the one where there is no violin at all. This view Blackburn has named the "universalist view of thought"\(^63\). And on this view, thoughts are not identity dependent on their objects, as one could have such a thought even in the absence of such an object. Blackburn says that "we should represent the psychology of the thinker best by referring directly only to the mode of presentation itself, rather than the object which it latches onto"\(^64\). Blackburn is adhering to a Russellian view of ordinary singular terms which are nothing but definite descriptions. And since definite descriptions are identity-independent and universal across context, so are these supposed singular terms. And as we have already said before, Russell has restricted the domain of singular thoughts to thoughts about oneself, about sense data, about one's own inner states. (We have discussed earlier how McDowell finds this Russellian restriction unacceptable as it leads to Cartesianism. Blackburn could be said to be committing the same mistakes by adopting this Cartesian stance. Blackburn has anticipated this charge and has tried to respond to it. We shall see how he does it later on in this section.)

We have seen from the first chapter that externalists would try to counter this kind of internalist thesis by saying that it is wrong to think that the thoughts of the subjects are same across the contexts. This is evident from the fact that if it is asked 'which violin does the subject

---

\(^62\) *Spreading the Word*, p. 313.

\(^63\) Ibid.
desire in \textit{w}_s^*\textit{,} then the answer would by Stradivarius\textit{1}, whereas the answer would be different if we asked \textquote{which violin does the subject want to buy in \textit{w}_s^*\textit{.} But Blackburn points out that though we describe the desires of the subject in the two different world differently (bringing in the relevant existing violin in that particular world), this does not indicate that the desires are themselves different. This thesis, which Blackburn calls the universalist thesis, would not be acceptable to the singular thought theorist, since the singular thought theorist refuses to accept that the same thoughts could be about two different objects and that there could be a thought without there being an object. A strong singular thought theory (SSTT) claims that the \textquote{subject characteristically has\textit{ expresses by his utterance no thoughts which are constant from world to world. He only has\textit{ expresses genuine singular thoughts.} Blackburn considers a weaker version of a singular thought theory (WSTT) which claims that the \textquote{subject has\textit{ expresses by his utterance universal thoughts. But in addition he has\textit{ expresses by his utterance genuinely singular thoughts\textsuperscript{65}.}

Blackburn finds the SSTT unacceptable on the grounds quite similar to the grounds that Segal has. Under this theory, there can be no empty singular thoughts. If someone \textquote{thinks\textit about a thing that does not exist, then she doesn\textquote{t think at all, under SSTT. But she might be acting on this \textquote{thought\textit of hers. And we might be able to give a psychological explanation of her behaviour, only if we take her to be thinking. Under SSTT there is no way of explaining actions of deluded subjects.

\textsuperscript{64} Ibid., p.314.
\textsuperscript{65} Ibid., 318.
The SSTT might come up with an argument against Blackburn. Under his universalist thesis, understanding lies in getting the mode of presentation of the speaker right. But I might get the mode of presentation right (i.e. understand the definite description she applies perfectly), and yet might end up misunderstanding her. Suppose I am speaking to you. And we both know Jerry and we both think that he has only one son. But in fact Jerry has two sons Paul and Erick. I know Paul (though maybe not by his name, but just as the only son of Jerry). And you know Erick just as the only son of Jerry. If I tell you 'I met Jerry's son in Battersea Park last Sunday', then in this case though you perfectly understand the mode of presentation associated here, yet you are bound to misunderstand me. You need to know which person I have in mind, otherwise you would be thinking that I saw Erick in Battersea Park, which I didn't. Nor did I want to convey that I saw Erick by my utterance. But against this the universalists would ask the same kind of question: if 'shared reference is the requirement permitting understanding in the normal case, what corresponds to it in the empty case?'

Here I would like to bring in an example from an Internet essay by Mark Textor entitled 'The Semantic Challenge to Russell's Principle'. He says in the essay that we need to individuate thoughts/propositional attitudes partly by the objects they are about for our ordinary explanations of behaviour. We find that we describe the actions by the use of singular terms. Suppose Harry marries Joan. And we want to explain why Harry married Joan. We could say that he married Joan because he thought (among other things) that Joan is beautiful. Textor says that while explaining Harry's action 'we cannot substitute "Joan" for

---

66 Ibid., p. 320.
a non-indexical definite description without the explanation loosing its force'. This is so because it is Joan that he marries and he marries her because he wants to marry her. And it is not as though that he would marry any one who fits the description 'the nicest blonde girl in Nottingham', though that is how Harry might be thinking of Joan and hence thinking that she was beautiful. When we explain Harry's action by citing his belief that Joan is beautiful, we would not be able to explain this action unless Joan existed. And so his action is explained at least partly by the object of his belief. May be this is too naïve an argument, as the relation between the explanand and the explanandum is far more stricter than that which is demonstrated here. The explanation in terms of the beliefs with the contents should be such that there is a strict relation between the two. If Harry married Joan because she is pretty and she is the nicest blond in Nottingham, then in the absence of such features in Joan Harry should not be marrying her. That is what is required of an explanand, if the relation has to be strict. We shall be discussing all these in the third chapter of this thesis.

An universalist thesis of the Russellian brand denies the kind of singular thoughts that the SSTT or the WSTT are subscribing to. They would say that suppose X has a thought to the effect that $a$ is $\phi$, where $a$ is a singular term. In such a case, $a$ has to be rightly related to X's thoughts. And in order to be so related, X's thoughts do not have to be intrinsically related to $a$. For the universalist, there may be two ways of describing a thought. One by mentioning the object (the de re style) and the other by not mentioning the object but by mentioning the way in which it is presented (the de dicto style). But there is just one kind of thought, which is thus describable in two different ways.
According to the universalists, we do not need to know the objects in the actual or substitute or empty possibilities in order to understand the subject's utterances in the actual, substitute and empty situations. They say that because 'we know what it would be like, from the inside, to be in the position of my subject in the various possible cases', we can see that the thoughts across the contexts would be the same, 'regardless of how they were causes'.

We have already said that this universalist thesis adheres to Cartesianism. On this thesis too, one assigns undue authority to the subject, whose knowledge of the contents of his own mind is made into something certain and incorrigible. Blackburn has tried to defend his thesis against the charge of Cartesianism. He says that all that he wanted to emphasise was that the twin earth experiments and his own thought experiments were legitimate. And they brought out that there was some common invariant element of thoughts across the context. This surely indicated that commonality was a legitimate category and hence the question whether we could bring all thoughts under this category too was legitimate. He also wants to emphasise that by saying that all the thoughts had this important common feature (which is how it feels to the subject), and it is this common feature that is important to us while we individuate thoughts and understand utterances, he is not making his thesis Cartesian. He says that if we look at the twin cases, we will see that it is not only the subjective states of the twins which are same. Even the environmental elements (like light and sound waves) which impinge upon the subject for it to have the experience are also same. Blackburn says 'one might even hold that psychology, and the theory of thought in a psychological sense, is concerned not only with what happens “in the

68 Ibid., p. 324.
head”, but also with certain relations between the subject and its environment, yet still oppose WSTT⁶⁹.

This response of Blackburn is however not acceptable at all. Let us see what exactly he is saying. He says that the charge of Cartesianism against him is redundant. This is so because he does consider the ‘relations between the subject and its environment’. But how important are these relations in his scheme. They are definitely not intrinsic to the thoughts of the subject. As the subject could have the same ‘how it feel’ property of the thought in the absence of such relations. Consider the case where the subject is hallucinating. Moreover, if Blackburn had taken these relations as intrinsic, then he would not oppose the WSTT. And if he does not make these relations intrinsic, then the kind of thesis he is adhering to is again open to the charge of Cartesianism. This is also clear from his open avowals to the Russell’s restriction.

What follows from our discussion of Blackburn is that Blackburn or any one who does not lift Russell’s restriction is bound to be Cartesian. In other words, all of those who would adhere to the Russellian restriction would assign incorrigibility and absolute authority to the subject about her inner states. And it is this inner character of mental states which is common across the worlds, which is a legitimate category under which we can bring in all mental states. We will therefore identify mental states in terms of this across-world-invariant-character they have. Now let us consider two states X and Y; let us see how we can determine whether these two states are the same or not. Under this theory, if X and Y have the same inner character, then they are the same. Now how are we to decide that they indeed have the same inner.

⁶⁹ *Spreading the Word*, p. 324.
character? We have to ask the subject, as it is the subject alone who has an access and an authority over these characters. And so if the subject says that her states X and Y have the same inner character, i.e. they seem to be same to her, then X and Y are the same. So we see that an internalist like Blackburn has to adhere to the first assumption, as he adheres to the Cartesian thesis.

We can now go back to Segal's criticism of SSTT. Perhaps what Segal finds unacceptable is that McDowell is suggesting that in the cases where the subject is deluded about the object, the subject is also deluded about the state of her own mind. This McDowell says, because he denies the possibility of there being empty singular thoughts. Segal says that this would lead to a serious problem. The problem that we will face is a problem regarding how actions of a subject who has been so deluded can be rationalised. But we do rationalise the actions of a deluded subject. For instance, if I were under the delusion that the door in front of me has an opening, the height of which is less than five feet, then I would bend my head. SSTT would not be able to explain my head bending behaviour in terms of any object-dependent thought. Suppose Ted was under the delusion that he was about to be attacked by a bear where there were no bear at all, then he like any other person who is actually being attacked by a bear would roll into a ball. Here too we cannot explain Ted's behaviour in terms of any object-dependent thought. But none the less, we can give a coherent explanation of his behaviour and such an explanation does not depend upon any object-dependent thought. Segal says that the SSTT cannot solve these problems without risking self-destruction. What is evident from this, Segal says, is that we can explain Ted's behaviour by other beliefs that he holds. And these beliefs are beliefs which his non-deluded counterpart Ed (who was really about to
be attacked by a bear and who behaves exactly in the same way as Ted) shares with him. If we can explain Ted’s behaviour with the help of those other beliefs which he shares with Ed, then we could equally explain Ed’s behaviour in terms of those very shared beliefs. So it seems possible that we explain both Ted and Ed’s behaviours in terms of thoughts which are not object-dependent. If explanation can run smoothly in one case without invoking object-dependent thoughts then it ought to run equally smoothly in the other case as well.

We see that Segal’s argument is designed to prove that it is wrong to think that the subject can be deluded about her own mental states. Consequently, we can say that Segal adheres to the thesis that the subject cannot be deluded about her own mental states. And so this thesis too is a Cartesian one. And like Blackburn, Segal too would be adopting the first assumption, viz. that if the subject cannot tell the difference between two mental states, then they have to be the same.

In spite of the Cartesian character of Segal’s argument against the externalist, this argument is a strong challenge against externalism. And I feel that this challenge has to be answered directly, otherwise the externalists would be in trouble. What I want to say is that it would not suffice to say, in answer to Segal’s charge (that if the externalists rule out the possibility of empty thoughts, then they would not be able to explain the actions of the subjects of empty thoughts), that his view is Cartesian. In order to answer Segal’s charge, externalists also have to show how the actions of deluded subjects can be explained within an externalist theory. Externalists cannot just discount this argument on the charge of its being Cartesian. I feel that we need to look into this argument and see if there
can be a direct way open to the externalists to meet the challenge posed by it.

Segal himself speaks of McDowell’s response to the charge that the SSTT cannot explain the actions of a deluded subject. We find in McDowell’s ‘The Sense and Reference of Proper Names’70, his suggested way of explaining the actions of a deluded subject inside the SSTT. He says that there is a significant difference between the explanation (or the interpretation) that we give of the actions of a deluded subject, and the explanation (or the interpretation) we give of the actions of a non-deluded subject. When we explain the actions of a subject, we usually do so by ascribing beliefs to her. In the case of the non-deluded person who utters let’s say ‘A bear is about to attack me’, we explain his utterance and his rolling into a ball behaviour by making a transparent71 ascription of the belief that a bear is about to attack him. We cannot make such a transparent ascription to Ted who is deluded that a bear is about to attack him. The kind of ascription we make to Ted is this: we say that Ted believes that he has a belief that can be expressed in the transparent mode to be a belief that a bear is about to attack him. It is this second order belief which is actually doing the explaining. And McDowell says that these two kinds of explanations are absolutely different.

Segal has come up with an argument against this answer of McDowell’s. If we look at the twin cases, then we will see that the second order belief which forms a part of other conditions suffices to explain

---

70 See Section VII of the paper, pp. 124-127.
71 Just as beliefs can be transparent and opaque (keeping in mind the transparent and opaque occurrence of a name in them), ascriptions of beliefs too can be transparent or opaque. Like here we can say that it is of the bear that Ed believes that it is about to attack him. Here we are bringing out the term ‘the bear’ outside the belief context and hence we are making a transparent ascription of the belief to Ed.
deluded Ted’s behaviour. This second order belief along with the other conditions are present even in the case of non-deluded Ed. And if this suffices to count as explanations of Ted’s behaviour, then, Segal thinks, they ought to suffice as explanations of the non-deluded Ed. We do not need to bring in the object-dependent belief in order to explain Ed’s behaviour either. As far as explanation of behaviours goes, this object-dependent belief remains idle.

And so Segal concludes that the ‘SSTT faces an unpleasant dilemma: either object-dependent thoughts outside Russell’s restriction are always explanatorily redundant, or sometimes the actions of the subjects who act on the basis of what they take to be singular thoughts are not rationally explicable’. It is clear from this that in order to avoid running the risk of facing this kind of dilemma, the externalists have to concede in some way that empty thoughts can come into explanation. A singular thought theorist might give up McDowell’s strong thesis that there cannot be empty singular thought, and still try to defend her externalism. If an externalist can show that in spite of the fact that there can be empty singular thoughts, the contents of those thoughts are quite different from the contents of their non-empty counterparts. But surely that content can be specified without bringing in any object. And hence there seems to be no reason for not accepting such an object-independent explanation in the case of a non-deluded subject, since such an explanation is available in that case as well. This kind of limited STT will actually have to say that the content of a deluded subject is different from the non-deluded one only as far as their reference is concerned, they are similar in every other respect. Such a two factor theory of content will not go too far, as this difference in reference would not be
reflected in any differences in the explanations of the actions of the deluded and the non-deluded subjects (an explanation which is made purely in terms of object-independent beliefs).

Segal considers another possible STT which he calls the LSTT$_2$. Under this theory, though all the ordinary singular senses are vulnerable to the kind of argument that Segal is bringing about, yet we could have Russellian logically proper senses which are not so vulnerable. This is because these senses are definitely object-dependent. Segal tries to show that these logically proper senses would fair no better. These are senses associated with demonstrative expressions like 'I', 'now', 'this sense datum'. We could consider the last one first. A sense datum is definitely not outside the subject's mind and so it could not be object-dependent. Now let us consider the other two. It is not very difficult to show how 'I', and 'now' actually have object-independent sense. And so such a LSTT$_2$ would fair no better.

I, however, feel that the way that the limited ST-theorists have tried to defend their thesis is not the right way of trying to save a singular thought theory from Segalian charges. I still feel that there is a solution nested in the argument that McDowell has come up with. In order to show that the arguments that Segal has brought against McDowell do not hold, we have to show that the second order beliefs of the deluded subject in terms of which we can explain his empty behaviour are some how linked to some object-dependent belief. Let us go back to the story of interpretation of behaviour. When I find Ted shouting out 'There's bear coming!' and rolling up into a ball, but cannot

---

73 Ibid., p. 47.
find any such bear in the vicinity, then why do I assume that he believes
that he has a belief which can be transparently described as a belief about
an attacking bear? Why do I not think of any other beliefs that Ted might
have, like *It is extremely hot today* or *Nick Cave has a wonderful voice*? We do
not, because of the following reasons. Usually when people think that
they have a belief about a belief that can be so transparently described,
they do indeed have this transparently describable belief. And this is so
because delusion is an exception, we usually are not deluded. And in
those cases where we are not deluded, it is in terms of the object which is
present that an interpreter understands and interprets our behaviour. If
people were deluded as a rule, and if people were usually wrong about
their second order belief, then it would not be possible for us to ascribe
this second order belief to Ted in this case. The very fact that we can
make distinctions between the deluded and the non-deluded, indicates
that we can understand that some beliefs are object involving. It is
because in most of the cases the second order belief is true, that we can
realise that in the deluded case it is false. But this is where the debate is
nested: whether or not this second order belief can be false at all, i.e.,
whether or not we can be mistaken about our own mental state. The
internalist would say 'No', while the externalist would say 'Yes'. And so
we again come back to the question about Cartesianism. And again we
come back to the question whether the assumption — that if a subject
cannot tell the difference between two mental states of hers, then there is
no difference — is right or not comes back. We come to the question
whether subjective indistinguishability is the most important thing to
consider when we are concerned with the individuation of mental states.
And this we will discuss now.
2.3 Identity and Indiscriminability

The concept of individuation is often used in various different ways. It is often spoken of as a purely metaphysical notion. One might say that in finding the individuative property of a thing we have to find the essential property of the thing and not the property by reference to which we/subjects manage to pick it out. But there are others who mean by individuative properties, the properties that help us, the subjects, to pick the thing out. So we have to be careful when we are speaking about individuation. We could be talking about two different things. On the one hand, when we are talking about individuation of mental states, we might be talking about a kind of a criterion that we need to use so that we can identify, or pick out mental states. On the other hand, we might be talking about something which constitutes the very identity of mental states. We could be talking about some kind of criterion of identity — which is more of a metaphysical concept rather than an epistemological one (which the former is). The metaphysical criterion is what the mental states have to satisfy in order that mental states constitute a category distinct from other metaphysical categories. This criterion is not something which helps us to distinguish mental states from other metaphysical categories. One might think that it is the handy criterion of individuation which we use, which has close connection with the kind of explanation that we want to give in psychology and the philosophy of mind. But the metaphysical criterion and the epistemological criterion are at least conceptually distinct, if not

---

75 Such views have been expressed by Kripke in his *Naming and Necessity*, Harvard Univ. Press, Cambridge, Mass, 1980 and by Putnam in his earlier writings (e.g. in his
more. When David Wiggins speaks about individuation in his *Sameness and Substance*\textsuperscript{76}, he says that to 'individuate' is to 'single out' (and this is in tune with the definition of 'individuation' found in the Oxford English dictionary). And to single out would mean to isolate the thing in experience by distinguishing it from all other thinks of like or unlike kinds. It is not always clear what this would involve. One could say that one can really single out a thing, say $x$, only if one hits at the properties which objectively belong to $x$ and $x$ alone. Alternatively, one could say that one can and does distinguish $x$ from others by means of the properties that strike the subject as uniquely belonging to $x$. Thus we see that the notion of individuation has been used in various different ways. We need to make this notion of individuation more precise, as the whole debate between externalists and the internalists is about how mental states are to be individuated. So in what sense do the externalists and the internalists take the notion of individuation? It seems to me that if the internalists are subscribing to (a), then they are speaking about the epistemological rather than the metaphysical criterion of identity. In other words, they are speaking about what it takes for us to know one mental state from another, rather than speaking about what it takes a state to be the same as or different from another one. And so for them if the subject cannot discriminate between two mental states, then the mental states have to be type identical. And if one accepts this, then one surely is opting for a view that (at least in the case of mental states) identity and indescribability coincide.

I have serious reservations about bringing the notions of identity and indiscriminability together. To start with, I would like to discuss

\textsuperscript{76} 'Meaning and Reference', in Schwartz (ed.), *Naming, Necessity, and Natural Kinds*, Cornell University Press, Ithaca, 1977.)
what the general problems are in taking identity and indiscriminability to coincide. The epistemology of identity, as we all know, is fraught with mysteries. What would knowledge of identity of anything consist in? Can we take indiscriminability as the mark of identity? This definitely seems to be what the internalists are subscribing to, at least as far as the question of the sameness and difference of mental states are concerned. In their argument against the externalists they are saying that since neither Ted nor Ed seems to detect the difference between water and twater as they do their world-hopping, their thoughts in both earth and twin earth should be type-identical. And having subscribed to the Cartesian picture they assign authority to the subject over her own mental states. They thus are taking indiscriminability of mental states as the mark of (type-)identity. If two mental states are indiscriminable to the subject, then they are type-identical.

But isn't discrimination a cognitive act? One might either succeed or fail in discriminating, depending upon one's own capacity to discriminate and the resources for discrimination available to one. But one's failure or success cannot decide if metaphysically there is a relation of identity between two states. It seems that if the internalists were to make the notion of identity too close to the notion of indiscriminability, then they would be robbing the notion of identity its own metaphysical character, as I shall demonstrate.

There are (1) general difficulties in bringing identity and indiscriminability together. But there are (2) specific difficulties, if
individualists try to say that in case of mental states identity and indiscriminability converge. Though it would be rather absurd to think that the internalists are unaware of the general difficulties in this matter, yet it would be a good idea to state what they are, in order to get a good grasp of the two concepts at play here, i.e. identity and indiscriminability.

2.3.1 General Difficulties

The basic difficulty in bringing the notions of identity and indiscrimination together is that, as metaphysical notions the two are very different. Saying that \(a\) and \(b\) are identical and saying that \(a\) and \(b\) are indiscriminable are not the same. Indiscriminability is an epistemic notion, while identity is not. Internalists, in order to tackle the general difficulty with indiscriminability, may try to give a non-epistemic or non-cognitive interpretation of the notion of discrimination. But it seems to me that such an interpretation of the notion of discrimination would be a sort of fancy 'metaphorical extension' (as Williamson puts it) of the concept. The concept of discrimination is indeed an epistemic or cognitive concept and it is natural to think that while explicating the notion of discrimination, we do need to keep room for failure of discrimination as well as success. The subjective failure of discrimination would not definitely be a mark of absence of identity. If we look at discrimination, this will be clear. What we are saying when we say that \(a\) and \(b\) are identical if they are indiscriminable to the subject in question, is this: \(a\) and \(b\) are identical because a subject cannot discriminate between them. There is a close connection between knowledge and discrimination, and it is perhaps this: to discriminate between \(a\) and \(b\) is
to know that \( a \) and \( b \) are distinct. One cannot discriminate between \( a \) and \( a \), as we can never know \( a \) to be distinct from \( a \). But when would we be able to vouch that we do have a knowledge that \( a \) and \( b \) are non-distinct? This is what the crucial point is. As the concept of indiscrimination is so closely tied with the concept of knowledge, we have to keep room for failure to discriminate as well. This is so because when one is using an epistemic concept, then one cannot make the concept intrinsically so conceived that it rules out the possibility of failures. All epistemological theories have to provide us not only with a theory which explains how our beliefs turn out to be true, but also with a theory which explains how error occurs. Epistemic concepts must allow for the possibility of error. But if we make the concept of discrimination a metaphysical concept, then we cannot speak of false discrimination at all.

We may take recourse to Leibniz's Law in order to discern the distinctness of \( a \) from \( b \), i.e. that identity entails sharing of properties and relations. So if there are some properties and relations that two things do not share, then they are distinct from each other. But we may just fail to notice some property and relation of an object at a particular time which we might later on manage to notice. This is the kind of thing that happens to Ed and Ted in the twin-earth example, when they get educated in chemistry and manage to discriminate their 'water' thoughts in earth from their 'twater' thoughts in twin-earth. Indiscriminability actually should be defined relative to time. For instance: \( a \) is indiscriminable from \( b \) for a subject at a time \( t \), if and only if at that time \( t \), the subject is not able to discriminate between \( a \) and \( b \) (i.e., if and only if at that time the subject is incapable of activating the relevant kind of knowledge that \( a \) and \( b \) are distinct).

---

We should bear in mind that discrimination is also an intensional notion. We know that knowledge is an intensional notion. Let us try to illustrate this point with the help of an example. I have two friends, one who looks like Mr. Bean and one who looks like John Lennon. Suppose you know this fact. Now the friend who looks like Mr. Bean is a space craft technologist, while the one who looks like John Lennon is an I.T. engineer. Given that you don't have this additional bit of information, you would be able to discriminate between the Mr. Bean look-alike and the Lennon look-alike, but not necessarily between the space craft technologist and the I.T. engineer. (More so because space craft scientists often are so well versed in I.T. matters.) So it is obvious that discrimination is an intensional notion.

There is another crucial difference between identity and discrimination. While identity is transitive, discrimination and indiscrimination are not. I shall take two of the examples that Williamson takes up to illustrate this point. Goldbach's Conjecture, being still a Conjecture, has not been either been established as true or as false. But it certainly is one of the two (according to classical logic). Now if we take, on the one hand, the truth-values of the statement '0 = 0' and truth-value of the statement '0 = 1', then each of them would be indiscriminable from the truth-value of the Goldbach's Conjecture, yet would certainly be discriminable from each other. Let us take another example, suppose you turn on the radio and hear three songs in succession. You have however missed the announcement and consequently the first singer is indiscriminable from the second singer and the second from the third. But the first is discriminable from the third and yet it might as well turn out that the second singer is in fact the
same person as one of the two other singers. All that these seem to suggest is that indiscriminability and identity do not always coincide. The cases demonstrated above are cases where there is indiscriminability but not identity. So indiscriminability cannot be a sufficient condition for identity.

It seems that one’s ability to discriminate \( a \) from \( b \) would be sensitive to the ways in which \( a \) and \( b \) are presented to the person. But the relation of identity holds between things, no matter how they are presented to one. For instance, my Mr. Bean look-alike friend is identical with himself and this is true no matter how one knows him. But whether or not he is indiscriminable from another person, say the Lennon look-alike, depends on how he or the Lennon look-alike is presented to the subject. Maybe you can discriminate him as the Mr. Bean look-alike friend of mine from the Lennon look-alike friend of mine, but cannot discriminate the Lennon look-alike from him as being my IT-engineer friend. Again, maybe you can discriminate the Mr. Bean look-alike from the Lennon look-alike as the one who makes paella (as you know that one of them is from Valencia and the other one comes from Toulouse, given that you know the connection between Valencia and paella and given that you know that people from Toulouse are not usually great paella makers). But you cannot discriminate one from the other as the friend of mine who juggles, because, though you don’t know, they actually both juggle. So it is clear from this story about my friends that whether or not you can discriminate one from the other depends largely upon what you know about them, and how they are presented to you. We must give an intensional reading of the notion of discrimination and indiscrimination. But such a notion is very different from the notion of identity, as identity is a relation between objects which can be detected
as distinct or the same by means of Leibniz' Laws. Perhaps one may say that we could give a non-intensional reading of the notion of discrimination. But that too seems to be yet another 'metaphorical extension'.

2.3.2 Specific Difficulties

It is really not plausible that the internalists are ignorant about the general difficulties in bringing the notions of identity and indiscrimination together. But they tend to suggest that though the notion of identity and indiscriminability are distinct, yet in the case of mental states the epistemic concept of indiscriminability and the metaphysical concept of identity go together. Now what exactly do they mean when they say that they go together? In a very simple and radical way they might mean that the two concepts converge in the case of mental states. If subject cannot distinguish between two mental states, they are identical. And this is so because following the Cartesian line, the subject is never mistaken about her own mental states. On the other hand, the internalists may not opt for such a radical thesis. They might say something less radical, viz., that in the case of mental states subjective indiscriminability can be regarded as a good evidence for the sameness of mental states. I would first like to rule out the first thesis which can be called the infallibility thesis. And then I shall try to show that the reasons for rejecting the first thesis may be regarded as good reasons for giving up the second one as well.
2.4 The Infallibility Thesis

Now we can go on to ask the real question: Is it correct to think that a subject is always aware of and correct about all the qualities of her experience? A Cartesian view, which assigns authority to the subject over all her mental states, would refuse to consider this as a possibility. And it is this which the externalists would question. We should, however, be careful about one thing: when we are speaking about a subject's being unaware of the properties of her mental states, we are not talking about unconscious mental states. We have said at the very outset, we are not concerned with unconscious mental states. We are concerned with mental states about which the subject is aware of and it is about these states the externalists claim that the subject could be mistaken. Of course, we have found in the history of psychology ample evidence in favour of unconscious mental states, cases where the subject is not aware of her mental states at all. This is definitely brought out in the case of unconscious mental states like unconscious hopes, desires and fears. For example, one might not know that one is expecting for a specific result in a match while one is watching it, but only realises it afterwards when one is disappointed over the result. Then one can hardly be expected to know the neurophysiological aspects of one's own experience. But we would not be concerned about all these kinds of mental states here. Again, sometimes it happens that the subject is unable to articulate the mental state she is in. When I speak of a subject's not being aware of all the properties of her mental states, I am not talking about this kind of failure of articulation on the part of the subject. Of course, one who has an inarticulate awareness at least can detect if some difference creeps in.
Whereas if a subject is unaware of all the properties of one of her mental states, then she could mistakenly take it to be identical with another mental state, not being able to detect that the difference that lies between the two mental states in terms of the property that she is not aware of. But the cases which I have been speaking of are cases where there is lack of awareness of certain properties of the mental states which marks the difference of one mental state from the other and not just lack of articulate awareness.

The reason that some philosophers have found it hard to accept that we can be mistaken about our own mental state is that there is one aspect of mental states about which we cannot be mistaken. And these philosophers have given more importance to this aspect of mental states than they have given the other aspect. There are two aspects of a mental state: the subjective and the objective. Traditionally the subjective bit has been regarded as what we might call the ‘how it feels’ character, or the phenomenal character. The feature that marks the subjective is that the existence of the subjective part of a mental state is dependent upon the existence of the mental state. The subjective bit of the mental state is what it is for the subject to be in that particular mental state. As Nagel puts in *Mortal Questions* for an organism to be conscious is ‘there being something it is like to be that organism’. The reason that no amount of reduction of this subjective element to anything else succeeds is that all that these subjective features of experience were reduced to were ‘logically compatible’ with the absence of these subjective features.

Whenever we try to understand the mental states of other persons, we

---

79 See Williamson’s, ‘Is Knowing a State of Mind?’ for a discussion on this.
81 In ‘What is it to be a bat?’, *Mortal Questions*, pg. 166.
82 For a detailed discussion on this see D. Chalmers’ *The Conscious Mind*, Chapters 3-5.
try to see them in the light of our own experiences. But no amount of imagination would help us to know how it is to be them or to have their experience. For instance, suppose while cooking pasta I put two chillies in the sauce. And at dinner I and my friend make the same mistake of not noticing them and biting them. Now in this case, in a way, I can understand how my friend felt when the hot chilli burnt his taste buds, as I felt the same. But still all that I would understand is that he felt something similar to me, but not how he felt it in his mouth as the burning of his mouth. And this is true even of my own experience over time. For instance, I don't know how my future experiences would feel like. For instance I don't know now, how it would feel like 'when I am sixty four'\(^3\). And since we have noted that the internalists have always said that after all what matters is how it appears to the subject, it might not be too wrong to think that the internalists put more importance to this private element of experience.

Both the externalists and the internalist can admit that there are these two aspects of a mental state: the subjective and the objective. They also agree that the objective bit of the mental state is what constitutes the content of the mental states. Now the real debate between the internalist and the externalist is regarding whether the content should be narrow or wide. The reason that internalists have opted for a narrow rather than a wide content view is that they give more importance to the subjective bit of the mental states, viz. the 'how it feels property'. That is, they seem to be thinking that we have to understand the objective aspect of the mental state in the light of the subjective aspect. This is so because (as I have discussed earlier) it is the 'how it feels' property which is invariant amongst all kinds of conscious mental

states. Suppose we were asked: What is this 'how it feels property' of a mental state? It will not do to say that it is the subjective property of the mental states. This is because the subjective property is definitionally the property of the mental state which does not depend upon anything outside the subject's mind. So in individuating this subjective property, we are actually bringing in narrow content. But it is to explain narrow content that we brought in the 'how it feels property'. And so in explaining the 'how it feels property', they cannot appeal to narrow content again. Whether or not one is opting for a narrow or wide notion of content depends upon whether or not a person puts more importance on the subjective or the objective aspect of the mental state. In settling this issue we have to decide which of the two would be considered as the deciding factor in connection with the question of sameness and difference of mental states. This is the moot point. And in deciding this we cannot in any way take recourse to the concepts of narrow and wide content again.

We must appreciate that there is one methodological point in favour of the internalists. In most cases if we cannot tell one thing from another, they are regarded as the same unless proven otherwise. This is something which the externalists too will appreciate. In discussing McDowell's response to the internalist charge that his theory cannot explain empty behaviour, we had mentioned this. The internalists have however extended this methodological wisdom into a strong metaphysical claim. The internalists have claimed that this methodological insight indicates an important metaphysical thesis: that in the case of mental states if the subject cannot discriminate between two of her mental states, then that shows that they are identical. In this case they claim that subjective indiscriminability and identity coincide.
The classic case where they do indeed coincide are usually what Williamson calls 'phenomenal characters', like looks, sounds, feels, tastes, smells, etc., i.e. appearances as they are called traditionally. Now when we are talking about the sameness and difference of mental states, is it the phenomenal character of mental states which is the most important thing to be considered?

We should approach this issue very carefully. Mental states and experiences are found out to be identical by taking recourse to the Leibnizian Law, i.e. by finding out if they have the same properties and relations. But the properties in question are of a very different kind from the usual ones. They are phenomenal characters. Now what are phenomenal characters? And why is it that in the case of phenomenal characters, if the subject cannot detect any distinction between them, then they are identical? The internalists could say that this is so as the subject can never be mistaken about the phenomenal characters of her mental state. And so, if she cannot distinguish between two mental states, they must be identical. If we agree that in the case of phenomenal characters the concepts of indiscriminability and identity converge, then we have to agree with the internalist position.

May be we should mention here that just to focus on phenomenal similarity does not presuppose independence from the world and hence internalism. It is only because the it is believed of the phenomenal characters/properties that the world had no role to play in their determination, that we can move from phenomenal similarity to internalism. We shall discuss in detail about phenomenal characters and their roles in the individuation of mental states shortly.
Having emphasised the subjective character of mental states internalists are more interested in the question: How do we judge the difference in our mental state from inside? For them the option of making the difference in terms of the precise identity or nature of the objects towards which the mental states are directed is not open. So we have to make a distinction between two mental states in terms of (let's say) the internal characters of the mental states in question. But naturally, we will be applying Leibnizian Laws and compare the characters of the mental states in question. These characters cannot be in any way external to the subject, as seen from the above discussion. These characters have to be purely subjective, and these are in fact the phenomenal characters.

Phenomenal characters are treated like particulars, and each is tied to a specific subject in whose life it forms an episode at a specific place and a specific time. These phenomenal characters are in this sense unrepeatable. What is meant here is that they are characters of experiences which are tied to specific times and once one has an experience with a phenomenal character, no amount of posterior knowledge about the object of the mental state will make any difference to the character. And, according to the internalists, these characters are the ones which are important in the individuation story. And that is why even when Ed and Ted would get to know about the chemical composition of water and twater, it would not change the nature of their prior mental states about water and twater. Whenever Ed and Ted are having an additional knowledge of the chemical composition of water and twater, what is changing is their understanding of the objective bit of their prior experience. They would maybe never characterise their
prior experience about water in case of Ed in twin-earth and experience of twater (or 'water' as in twin-earth they call twater 'water') in the case of Ted. Their posterior knowledge about the chemical constitution of water and twater would not change the 'how it feels' character or phenomenal character of their experience.

The internalist line of argument seems to suggest that in order to detect the sameness and difference of some mental states (the conscious ones), we have to rely on their phenomenal characters alone. And in cases of phenomenal characters it is subjective indiscriminability that matters most. It is here where I find it difficult to accept the internalist thesis. The idea of phenomenal characters is such that it is actually definitionally impossible for a subject to be mistaken about the phenomenal character of her mental states. It is as though we are defining phenomenal characters as those characters of the mental states about which the subject cannot be mistaken. And then we are going on to say that in the individuation of mental states we have to depend on these phenomenal characters and so we can never be mistaken about our own mental states. Of course this would follow trivially from the very definition of phenomenal characters. But the real question is why should we put so much emphasis on the phenomenal character of mental states. As we have already said, there is another aspect of mental states which is not so subjective. Why can't we emphasise the objective aspect while individuating mental states? That is the real question. Unless the internalists give a good answer to this question, they would be begging the question by emphasising the importance of phenomenal characters in the individuation story of mental states. It is whether or not the world and the objects in it do have a role to play in the individuation of mental states, what we are trying to find out. And in doing so we cannot already
assume a characterisation of mental states in terms of phenomenal characters. This would be too arbitrary. We have to give good reasons for assuming the importance of phenomenal characters. If what decides the issue about sameness and difference of mental states is how it seems to the subject, then it would follow trivially that the differences in the external world makes no difference, unless they seem different to the subject.

To emphasise the importance of phenomenal characters in the individuation of mental states the internalists might have two important things to say. First of all, the internalists have continuously emphasised that there is something which is invariant to all mental states, no matter how the world in which the subject is, is like. This invariant element of all mental states is the phenomenal character. We find in philosophical discussions that when we want to individuate a category, we try to find out an invariant character which is present in all the entities which belong to that category. In this case it is the category of mental states which we are trying to individuate. And so it would not be too outrageous to try to find something which is invariant to all mental states. And it is indeed the phenomenal character which is invariant to all mental states and hence it is not outrageous to give more importance to this character in deciding the sameness and difference of mental states. This could well have been the reason the internalist had in favour of the importance of phenomenal characters in the individuation of mental states. We have seen similar arguments in the philosophy of perception. And we shall discuss briefly those arguments to show if such
arguments can take us far in philosophy (in general). We find Simon Blackburn mentioning this in his *Spreading the Word*.

The other reason that the internalists have for taking phenomenal characters as important in the individuation of mental states is this: If we do not identify mental states in terms of such a character about which the subject cannot be mistaken, then we would not be able to account for the possibility of self-knowledge. We usually believe in first person authority. If we emphasise on the character of mental states about which a subject can be mistaken, then we would be questioning the possibility of first person authority.

We shall discuss these two reasons separately in two sections\(^{85}\) and I shall try to demonstrate that these two reasons might not be good enough to justify the emphasis that internalists have given to phenomenal characters. But there is another reason for which one can have doubts about this misplaced emphasis on phenomenal characters. And that is due to suspicions about phenomenal characters as such. There may be some serious problems relating to the concept of phenomenal character itself and so this could be a further reason for not accepting the infallibility thesis. What is important here is that one has to take up the phenomenal character talk, if one has to hold on to the infallibility thesis. And if we can manage to show that there is something wrong about the whole idea of phenomenal character or the way in which it gets used in the infallibility thesis, then we will manage to counter the infallibility thesis.

---

\(^{84}\) See p. 311, p. 317.

\(^{85}\) The common element thesis as it is found in the philosophy of perception and philosophy of mind will be discussed in the Section 2.6 and the problems regarding self knowledge will be discussed in the Section 2.7.
We shall first discuss the problems that we have regarding phenomenal characters and then we shall go over to the two other reasons that internalists have for understanding mental states in terms of phenomenal characters.

2.5 The Problems with Phenomenal Character

We find in the literature many philosophers having reservations about phenomenal characters. The problems that we have with phenomenal characters are mostly due to problems with the concept of a subjective kind. We also find problems with what is known as the paradoxes of phenomenal characters. For instance, Tim Williamson speaks about the various difficulties in the very concept of such subjective kinds in the fourth chapter of his *Identity and Discrimination*. But the main problem that Williamson thinks that these qualities would have is that if we have to make identity and indiscriminability coincide in the case of these qualities/phenomenal characters, then we have to say that they do not have one of their essential characteristics, which is presentation sensitivity.

The biggest problem with phenomenal characters is that it is impossible to make phenomenal characters transitive. In order to ensure that the subject has an infallible access to her own mental states, philosophers have brought forward the concept of phenomenal characters as unique and non-repeatable. But if this be the case, then can
we make sense of the different presentations of the same character at all? And if we cannot make sense of this, then can we make sense of two mental states having similar phenomenal characters. Two experiences might have the same character, and might also present this character in the same mode. But what is important to realise is that even then it would be distinct presentations of the same character. The things which have the characters or qualities are the instances of those qualities. To get a grip on this problem we have to know what it is for a character to be presented by an experience. It is important to know this because what we have to show is that each phenomenal character can be presented in just one way. If this is how phenomenal characters are conceived, then it makes no sense of talking about two presentations of the same phenomenal character.

And in a way there is no way in which the internalists can give up this conception of phenomenal character, because regarding such a character the authority of the subject would be incorrigible. But can we have such a conception? The kind of conception of phenomenal character that the internalists are opting for seems to be very close to (or even the same as) what is known as ‘qualia’ in the philosophy of mind. And so in questioning phenomenal characters I would like to bring in the concept of qualia, which has gained so much interest amongst the philosophers. Qualia are phenomenal characters. Now there are philosophers who are suspicious about such things as qualia\textsuperscript{86}. There are philosophers who think that it is not necessary at all to bring in the notion of qualia in order

\textsuperscript{86} For example Williamson in his \textit{Identity and Discrimination} and Dummett in his ‘Wang’s Paradox’ in \textit{Truth and Other Enigmas}, Harvard University Press, Cambridge, Mass, 1978, as we will discuss soon.

to understand the subjective aspect of mental states/experiences. And there are other philosophers who say that even if we grant that there is a need to posit qualia in order to make sense of the subjective aspect of mental states, yet the very concept of qualia is mysterious.

So we shall divide our discussion on phenomenal characters into two sub-sections. In the first we shall discuss the paradoxes of phenomenal characters as discussed by Dummett and Williamson. In the second we shall discuss about problems with the concept of qualia.

2.5.1 The Paradoxes of Phenomenal Character

There seems to be enough logical problems with the concept of a phenomenal character. The problem that the notion of phenomenal characters faces is that the non-transitivity of indiscriminability leads to Sorites paradoxes. This has been argued by Michael Dummett in his article ‘Wang’s Paradox’. He starts by showing that observation predicates are vague and in case of observable qualities or phenomenal qualities their non-discriminable difference too is ‘non-transitive’. If we take the example of observable predicates, like colour predicates, then this will be evident. We shall see that the same kind of thing happens in the case of observational predicates and observational qualities as we noted in the example of the songs on the radio in Section 2.3.1. Observational predicates are predicates whose application can be decided on the basis of the employment of sense organs. The problem with vague predicates like the phenomenal characters like colours is that if we try to make indiscriminability of them transitive, then we might

---

land into a paradoxical situation. There are cases where we have problems in telling an orange thing from a red thing. Suppose we have in front of us things of all different shades of red getting lighter and lighter fading into an orange thing. In such a case, it will be difficult for us to tell one colour from the one just next to it. But if we make indiscriminability transitive, then we should not be able to tell the first red thing from the last orange thing. This however is not what happens. We can tell the first as red and the last as orange. In the case of such vague phenomenal characters we know that if the subject cannot tell the difference, then they should be the same (as that is the very essence of phenomenal characters, viz. that the subject has a privileged and incorrigible access to them). So in the case where I cannot make a difference between the colour of $a$ and $b$ and I recognise $a$ as red, $b$ must be red too. This is so because in the case of phenomenal characters all that we need to do to recognise them as some kind is to look at them. But in case of the red and orange things, we cannot conclude just from the fact that we cannot discriminate between $a$ and $b$ (if we want to make indiscriminability transitive in this case), that if one is red the other has to be red as well. This is precisely because the 'non-discriminable difference fails to be transitive' in the case of vague phenomenal characters. And hence Dummett concludes:

1. If the 'non-discriminable difference' of an observational predicate is non-transitive, then that predicate has got to be vague.

2. In such a case the very use of such a predicate is intrinsically inconsistent.

And the reason that Dummett says that they are inconsistent is that ‘it would be impossible to force someone, by appeal to rules of use that he

---

acknowledges as correct, to contradict himself over whether the predicate applied to a given object.\footnote{Dummett, 'Wang's Paradox', p. 264.}

This shows that if the internalists rely on phenomenal characters in order to decide the identity and difference of mental states, then they will have a problem.\footnote{I am not suggesting that externalists do not make use of vague predicates and the vague properties corresponding to them, but since they do not make them essential to the idea of the identity and difference of mental states they would not have this problem.}

\subsection*{2.5.2 Problems with Qualia}

Our discussion of the phenomenal character of mental states and its importance in the question of individuation of mental states bring us to the heart of one of the main problems in the philosophy of mind. Philosophers have been trying to bring out the essential mental character in mental states since the beginning of our concerns with the mind. In a very trivial sense we wanted to bring out the non-physical character in mind which made it so distinct. Descartes had thought that this non-physical character of the mind was nested in its conscious character. And hence we find John Searl saying in his \textit{The Rediscovery of The Mind},\footnote{MIT Press, Cambridge, Mass., 1992.} 'The study of mind is a study of consciousness'. But what is this consciousness? There is of course a tendency towards finding a single property or character which is essential to consciousness as such. We have already noted how difficult it is to bring all kinds of mental phenomena under one single characterisation. 'What single property makes thoughts, feelings, sensations, emotions, and desires all mental?'\footnote{See Rosenthal's \textit{The Nature of Mind}, Oxford University Press, New York, 1991, p. 289.}
It would not suffice to say that all these states are non-physical and are hence characterised by consciousness. We have to say what this conscious character consists in. We find that even common people make distinctions between different kinds of consciousness: self-consciousness, intentional consciousness, phenomenal consciousness, introspective consciousness (though maybe they do not articulate the distinctions quite in this way). Which of these brings out the essential character of consciousness? Perhaps it is most natural to bring in the intentional and the sensory characters of the mental phenomena to understand the conscious character of it. And hence we find philosophers speaking more about phenomenal and intentional consciousness. But which of these two gets preference in the individuation story? This is the moot question. We know that for some philosophers it is the intentional character which is the most significantly mental character; we remember Brentano's dictum: intentionality is the mark of the mental. A mental state has this intentional character because a mental state can direct itself onto an object. Knowledge, beliefs, desires, intentions, hopes, fears, wishes are all intentional states. Intentionality is nested in memory, perception and intentional actions as well. Russell had called intentional states propositional attitudes, since they had as their content some proposition. And so in an intentional state we find that there are two elements: the attitude and the content of the attitude. And it is about this content that we are discussing: whether it is narrow or wide. We have seen from our earlier discussions that those who opt for a narrow content consider the phenomenal character of the mind to be more important. Now what is phenomenal consciousness or the phenomenal character of mental states?

See Amita Chatterjee's, 'The Intentional and The Phenomenal', paper presented at seminar on 'Consciousness and the Phenomenal Mind', Lucknow, India, February,
In current philosophical literature the phenomenal character of mental states has been identified with sensory qualia. Qualia are the subjective, private, experiential properties of mental states. They are the so-called 'raw feels'. If we make the phenomenal character of the mental states the individuative character of the mental states, then the subject can never be mistaken about it. And this is why the internalists have emphasised the phenomenal character of mental states, as opposed to the intentional character of them.

On the face of it, it might seem strange as to why any one should have any reservations about qualia, the subjective, private experiential properties of mental states. Now the reservations regarding qualia are too many and it is a topic which needs careful discussion. I do not intend to embark upon a long and detailed discussion of it here. All that I want to point out is that this is a concept which is not obviously acceptable. And I shall briefly state the reasons why people have found this concept suspicious. But before that I would like to say why some philosophers think that the subjective aspect of mental states/experiences can be explained without having to postulate such things as qualia.

In 'Visual Qualia and Visual Content', Michael Tye has tried to show that the subjective element of visual experience (and he says that this can be shown to be true of all other kinds of mental states) are determined by 'aspects of its representational content'. Tye says that the only reason that one might have to postulate visual qualia is that they explain the subjective or phenomenal aspect of visual experience.

1999.

9See Tye, 'Visual Qualia and Visual Content, p. 160
i.e. why the experience has these aspects. But in fact they do not explain them. If we postulate qualia in order to explain the 'felt' aspect of visual experience (as qualia are non-physical and irreducible things like felt aspect of blue), then having such an experience would be 'a matter of its having a special, non-physical property'. Tye feels that this is not a very enlightening story about our felt aspect of visual experience. It would not, for example, help us to understand why the felt quality of the experience of blue is not the same as the felt quality of the experience of red. (To say nothing about the mystery centring around the emergence and the causal role of such properties in the world.) But our introspective awareness tells us that the visual experience that represents blue is different from the one that represents red. And we have to make sense of this difference. According to Tye, it is the content alone which is able to explain this.96

Even if we accept that we need qualia to explain the subjective aspect of experience, there are serious problems with the very concept of qualia. The problem is that nothing is known about the brain which could explain qualia. And maybe the way that the opponents conceive of qualia make them irreducible in the intentional or functional or cognitive terms. For Dennett qualia, if they existed at all, were non-relational, incorrigible and have no scientific nature at all. For Nagel97 it is what is subjective par excellence. And as the domain of the physical is the 'domain of the objective par excellence', there is no way in which we might understand the qualia in physicalistic terms. McGinn calls this the

96 See the second section of Tye's article, where he demonstrates how this difference can be made only in terms of content and that the felt aspect cannot be divorced from the representational aspect of experience.
97 Mortal Questions, p. 172.
inaccessible nature of consciousness. In a way this subjective point of view is really what is singular to a mental state, because if you take away this, there would be very little left to the mental states. But the question is how to keep this aspect of mental states without making them totally mysterious and without opting for an incorrigibility thesis about them.

The main problem with qualia (the raw feels like the sweet taste of honey, or the hurtfulness of a tooth-ache) is a problem that is known in philosophy as the problem of the explanatory gap. There seems to exist an explanatory gap between facts about our physical constitution and facts about the phenomenal character of our felt experiences. Why do we, who are organisms constituted entirely by material particles, have a phenomenal character emerging in our experiences? How can a physical system like a brain ever manage to be endowed with a purely subjective awareness of things like colours, tastes, sounds, pleasures and pains? Experience, having such a phenomenal character, becomes a strange subjective feature of an apparently objective physical system. How can a purely physical thing have a purely subjective inner life? Qualia seem to be posing a problem for physicalism and functionalism. We find three kinds of arguments that are brought to show how qualia so pose a problem: knowledge argument, explanatory gap argument, and the inverted or absent qualia argument. It is not possible to discuss all these, here but I shall briefly state what all these arguments try to show.

---

98 See McGinn, The Problem of Consciousness.
Before we discuss these arguments I would like to mention that in recent years David Chalmers has spoken of the real 'hard problem'\(^{100}\) of consciousness. It is the question of how there can be a supervenience of subjective awareness on the objective physical or neuronal substrate. This needs serious consideration which again I shall not do in this thesis. But I would like to point out that we shall see later in the thesis that internalists have expressed worries about the idea of wide content since it cannot be accommodated with the idea of psychophysical supervenience. I would like to point out that if narrow content depends heavily on the idea of phenomenal consciousness, then it too would pose a problem for accounting psychophysical supervenience. At least it is not so obvious as to how something non-physical *par excellence* be supervenient on something thoroughly physical. If qualia pose a problem for physicalism, in the sense that they cannot be accommodated within a physicalistic framework, then I feel that an internalist thesis could have very little use of such a notion.

We can now go over to the discussion of the three kinds of arguments that demonstrate that qualia pose a problem for functionalism and physicalism. The knowledge argument as we find it in Frank Jackson’s famous article ‘What Mary Didn’t Know’\(^{101}\) tries to show that no amount of knowledge about the physical states accompanying the phenomenal experience would make one know what the phenomenal experience is like. Jackson speaks of a hypothetical case of a super colour scientist Mary who is brought up in a black and white environment, and knows every bit of physical fact that one can know

---


\(^{101}\) *Journal of Philosophy*, 83:5, 1986.
about colour vision and yet has no knowledge whatsoever about what it is like to see colour. This is demonstrated by the fact that as soon as she is brought into a coloured environment, she will realise what she was missing out on. This shows that Mary in fact did not know everything that there was to know about colour perception by merely knowing all the physical facts about it. This shows that there is more to colour perception than just the physical happenings associated with it. So Jackson's argument runs like this:

1. Mary knew everything physical about colour perception before she was released from her black and white environment.
2. Mary comes to know something new about colour perception after she is released from her black and white environment.
3. Therefore, there are some facts about colour perception which is missed by a physicalist theory.
4. Therefore, physicalism has to be wrong as it fails to explain colour perception.

This argument has indeed posed a very serious threat to physicalism.

Now of course there have been several attempts to show that this knowledge argument does not pose any serious threat to physicalism and that the notion of qualia can be accommodated in a physicalist theory. I shall briefly state the arguments that have been brought against the knowledge argument of Jackson. I shall take help of Robert Van Gulick's classification of the critical objections that have been raised against the knowledge argument. The main line of argument against the knowledge argument is that it is wrong to suppose that Mary the

---

102 See his 'Understanding the Phenomenal Mind: Are we all just Armadillos? Part I', Figure 35.1.
colour scientist does in fact gain some new knowledge when she experiences colour. Paul Churchland in his 'Reduction, Qualia and the Direct Introspection of Brain States'^103, refuses to accept that Mary knows something new when she comes to have experience of colour. This is so because our assumption in the argument was that Mary knew everything physical that there is to know about colour vision, and since our present knowledge of the brain is not complete, we cannot claim to have an idea about Mary's knowledge. And so Churchland concludes that the knowledge argument cannot prove physicalism to be false. But surely we do feel that Mary did come to know something new, and most philosophers will think the same. But the real question is: What kind of knowledge does Mary come to have? Some philosophers have tried to argue that the kind of knowledge that Mary comes to acquire is nothing but know-how and not knowledge of facts or propositions. She acquires just some new abilities which help her to recognise phenomenal properties. This is a view that has been subscribed by Lawrence Nemirow^104, and David Lewis^105. And so even on this view physicalism is not refuted by the knowledge argument. Some philosophers like Michael Tye^106 have however agreed that Mary's knowledge is not just a know-how, but it is indeed propositional knowledge. But the fact is that Mary does not acquire the knowledge of any new facts or propositions. She just knows old propositions in a new way. Philosophers like Brian Loar^107 have tried to show that even if we agree that Mary has come to know new propositions, this would be no threat to physicalism. He

argues that what we find in Mary’s case is acquisition of a new concept. This concept comes in with her new discriminative abilities. And all this helps her to acquire knowledge of new propositions. But this would not be a threat to physicalism, in so far as the property which her new concept refers to is the same property which her former and purely physical concept referred to. I shall not go into a detailed evaluation of these arguments. All that is clear from the debate centring around qualia is that it is not quite obvious how it could be accommodated within a physicalistic theory.\(^{108}\)

The explanatory gap argument says that the way the qualia are conceived (as private, paradigmatically subjective, basic simples) it is impossible to establish strict links between them and their objective physical or neural substrate. There have been attempts to bridge the gap between qualia, the felt quality of experience and the aggregation of millions of individually insentient neurones which make us up. But there is a serious worry about whether such attempts would end up in robbing qualia of their essential subjective incorrigible character. And for our purpose such attempts would be of no use since such qualia would be of no use to the internalist.

The absent qualia argument shows that human beings and robots can be computationally identical, and whereas there is something like being human, there can be nothing like being a robot. A similar argument is the inverted spectrum argument. This argument speaks of a hypothetical possibility where when you see green, you have the brain states which are identical with the brain states that I have when I see red.

\(^{108}\) See Van Gulick’s ‘Understanding the Phenomenal Mind, Part 1’ for a detailed discussion on the knowledge argument.
However you call your ‘green experience’ an experience of red, and vice versa. And so our experiences are inverted. Hence we see that we have everything else similar and yet we cannot be said to have the same phenomenal experience.

All these arguments demonstrate the problems that physicalism would have with the notion of qualia. We shall see in the following chapters that the internalists prefer narrow content to wide content, as wide content cannot be accommodated in the physicalistic framework which is essential for scientific psychology. If this is be so, then if we understand narrow content in terms of qualia (which might not be explainable in physicalistic terms), then we could not have a scientific psychology of mental states which are individuated by such a narrow content. This is the main threat that the idea of a purely subjective, private, incorrigible qualia might pose to internalism. In other words, the big problem with qualia is that it is difficult to understand how they could have relations with the physical world.¹⁰⁹

From our discussion of the problems that we have with qualia we now are again faced with the question: Is it the subjective aspect of mental states or the objective aspect of mental states which characterise the mentality of the mental states? Perhaps we could mention an alternative view here which considers consciousness a cluster concept. This view is based on the insight that both pure intentionality and pure phenomenality are problematic and hence we could possibly think of the mental as a blend of the phenomenal and the intentional. I shall not

¹⁰⁹ Subscribing to qualia would create problems in understanding mental causation. I shall be discussing the problem of mental causation in the fourth chapter as it has been usually seen as a problem that externalism specially faces. See Ned Block on ‘Qualia’ in
discuss this here but certainly this could be an alternative worth considering\textsuperscript{110}.

2.6 The Common Element View: Philosophy of Mind and Philosophy of Perception

We shall now discuss the intuitively appealing idea that in order to individuate a category we need to look for an element which is common to all the entities which belong to that category. This is an idea which might have led internalists to emphasise the phenomenal character of mental states as they seem to be common to all mental states. This view of the internalists resonates a kind of view that can be extracted from the arguments that we get in the literature on perception. And of course this is only natural as perception is a special case of mental states.

Recall the Argument from Illusion. In this argument what is being shown is that since phenomenologically, the putative veridical perception and the non-veridical perception are the same, so the kind of object perceived in one case is the same as the one that is perceived in the other. But that view has faced criticism from philosophers\textsuperscript{111}. And I want to take the line they take in questioning the Argument from Illusion, in dealing with internalism about the mental. Moreover, as we have seen

\textsuperscript{110} See Amita Chatterjee’s ‘The Intentional and the Phenomenal’ and Norton Nelkin on ‘Subjectivity’, in \textit{A Companion to the Philosophy of Mind}.

already, the very concept of subjective qualities, or phenomenal characters is a suspect.

Let us see first how philosophers argue in the case of perception and see if a parallel argument\textsuperscript{112} can be given against the internalists.\textsuperscript{113} Both internalists and philosophers who accept the conclusion of the argument from illusion are appealing to the phenomenal/phenomenological character of experience. In the case of perception, because the putative veridical and non-veridical cases are phenomenologically indistinguishable, the Argument from Illusion is supposed to show that the kind of object that is perceived in one case has got to be the same in the other case. In the case of mental states, since Ted's water thoughts and twater thoughts, across worlds, are indistinguishable, water or twater play no individuative role in his thoughts and hence his thoughts across the worlds are type identical. Just as in the case of mental state, in the case of perceptual experience there are two things that should be distinguished: the subject's capacity with respect to being able to discriminate the objects of experience, and the subject's capacity to discriminate one of her mental states from others. What is being suggested by the argument from illusion is that the objects towards which two mental states are directed might be different and the subject cannot detect it, but they rule out the possibility of the subject's failure to detect the difference between two of her own mental states. It is here that the arguments found in internalism and the arguments found in the literature on perception run on similar lines. The striking similarity between the two kinds of arguments is actually due to

\textsuperscript{112} Note that what these arguments address are a special case of the externalism/internalism issue.

\textsuperscript{113} See Michael Martin's article 'The Reality of Appearance', paper presented at a conference on 'Thought and Ontology' in Genoa, 1995.
the prevailing faith in the common element thesis in both the question of individuation of mental states, and the question of understanding of perceptual experience.

I will be discussing the common element thesis and its close connection with the subjective indistinguishability principle as we find it in the literature on the problem of perception. In this I shall be taking help of arguments that Mike Martin has given in his paper ‘Reality of Appearances’. In this article Martin is mainly concerned with perceptual experiences, but what he says can be extended to mental states as well. He too points out that there is a principle which is found by many philosophers to be extremely plausible, viz.:

‘(IND) If two perceptual experiences are indistinguishable for the subject of them, then the two experiences are of the same conscious character.’

Note that this principle (IND) is very similar to the presupposition that I claim the internalists have. The presupposition as I have formulated it, in the beginning of this chapter (to recapitulate) is:

We can consider two mental states to be the same (i.e. type-identical), if the subject cannot discriminate them. (Let's call this principle, the internalist's principle or the INTP.)

There is of course a noticeable difference in these two principles, besides the difference that IND is about perceptual experiences while INTP is about mental states in general. But this difference is not so interesting. What is important to us is that IND says that subjective indistinguishability is the mark of sameness of the conscious character of perceptual experiences, whereas INTP says that subjective
indistinguishability is the mark of sameness of mental states (simpliciter). So one might take these two principles to be different. But in fact they are not so different. If we remember that the internalists were looking for a common element invariant to all mental states and that they stumbled upon the conscious (phenomenal) character of mental states, we will realise why the INTP would be nothing but a version of IND.

What I am interested in finding out here is whether the kind of arguments that Martin has brought against IND can also be brought against INTP. With this query in mind, I shall be discussing Martin’s paper in brief.

Let us see the different kinds of responses that we find to the argument from illusion in the literature on perception and try to find if we can trace parallel theories in the internalist-externalist debate. To start with, we have the view which is radically opposed to the argument from illusion: the naïve realist view. This view, like all the other views, does recognise the phenomenal character of experience. But they, unlike others, think that the actual objects of perceptual experiences like birds, tress, clouds and rivers, and the properties that these external objects manifest when they are perceived, are what constitute our perceptual experiences (at least partly). And these are what determine the phenomenal character of our perceptual experience as well. ‘This talk of constitution and determination should be taken literally; and a consequence of it is that one could not be having the very experience one has, were the objects perceived not to exist, or were they to lack the

---

features they are perceived to have. Characterised like this, the naïve realist position is one close to the strong singular thought theory and the thesis regarding de re thoughts which we have discussed earlier. The kind of arguments that were brought against the SSTT is also brought against naïve realism. We found internalists arguing that a SSTT will never be able to explain empty singular thoughts. We now find similar arguments against the naïve realists which aim at showing that under the naïve realist thesis we cannot make sense of hallucinatory and illusory perceptual experiences.

There can be two alternative theories to naïve realism: the sense data theory and the intentional theory. The sense data theory is a subjectivist thesis which says that the perceptual experiences consists in the awareness of an entity whose existence and nature is dependent on the awareness itself. These are sense data which are subjective entities. The phenomenal characters of our perceptual experience are determined by these entities and their properties. Philosophers have found this idea of a subjective entity extremely problematic and so this is not a view that is accepted by many. The other alternative to the naïve realist thesis is the intentionalist theory of perception. The intentional theory of perception claims that the phenomenal character of an experience is exhausted by its intentional content. It is important to make clear whether intentionalism about perception is consistent with an externalist view of content, or an internalist view. This thesis, however, entails neither externalism nor internalism. We find an externalist philosopher like Michael Tye endorsing an intentional theory of perception, and also an internalist philosopher like Tim Crane endorsing to an intentional theory of perception. The intentional theorists agree with the naïve

\[115\] Ibid., p.3.
realist that the perceptual experience is about the external object, but they reject the naïve realist’s belief that the object is a part of ‘the relational state of affairs that comprises perceptual experience’\textsuperscript{116}. This thesis takes the objects of perceptions to be before the mind in perceptual experience, but they also say that these objects need not be so in order that one have such an experience. Of course we can find internalist line of thinking in the intentionalist theory. The internalists may well accept the object-involving nature of mental states but would refuse to accept that they strictly constitute the mental states, as one could be in the mental states even if the objects did not exist.

As is clear from the formulation of the thesis, the intentional theory is well-equipped to handle the hallucinatory and illusory perceptual experiences. Just as internalists claim that they can account for empty thoughts and explain for empty behaviours because they adopt narrow rather than wide contents of mental states. Under the intentional theory, object along with their properties can ‘be before the mind’ without actually existing.

We see that the intentionalists, as well as the sense data theorists, appeal to the common element thesis in order to combat the problem of illusion and hallucination, just as the internalist do in order to combat the problem of empty thoughts and empty behaviours. A naïve realist might retort to these kinds of theories by saying that the common element thesis is wrong. This can be found in John McDowell’s ‘Criteria, Defeasibility and Knowledge’ where he says that it is wrong to think that the object of experience in the veridical case is just like the object of the experience in the non-veridical ones, mere appearances. This he says

\textsuperscript{116} Ibid., p. 4-5.
because he denies that appearances have to be the intermediaries in all kinds of experiences. At least in the veridical cases it is not the appearances which are 'intervening between the experiencing subject and the world'\textsuperscript{117}. Arguing in this way the naïve realist would be actually denying the common element view altogether.

There is, however, a mid-path in between the adoption and denial of the common element thesis. This is found in the disjunctive theories of perception. We must mention at the very outset that the intentionalists need not endorse a disjunctivist thesis. This theory in fact questions the assumption that is shared by both the intentional and the sense data theories, which is 'that experience forms a common kind among perception, illusion and hallucination'\textsuperscript{118}. We can be thoroughly non-committal about which of the three kinds of perceptual experiences we were having. And when we talk about how things look to us when we experience them, we are not talking about the state of mind common to the three cases. Consequently we need not give the same kind of account of all the three kinds of perceptual experiences. The disjunctivist thesis claims that when a person has a perceptual experience of say a tree, since it looks to her as though there is a tree in front of her, then either there is indeed a tree before her which looks to her to be so, or it merely looks to her that there is a tree. We have to remember that the disjunctivists are talking about exclusive disjunctions. They are not saying that the disjunction can be made true by the fact that both the disjunctive cases have something in common.

\textsuperscript{117} McDowell, 'Criteria, Defeasibility and Knowledge', p. 211.
\textsuperscript{118} See Mike Martin, 'Perceptual Content' in \textit{A Companion to the Philosophy of Mind}. 
The disjunctive thesis might not be obviously helpful to the naïve realist thesis. This is so because the disjunctivist alternative definitely takes the delusive case as a distinct state of affair, and hence poses a problem for the naïve realists as they have no way of accounting for such a state of affair. The point that the naïve realists have to make is that the account of non-delusive experience has got to be distinct from the account of the delusive case. We find this kind of a view in Tim Williamson's 'Is Knowing a State of Mind'? He takes up an analogous case to demonstrate that the veridical case has to be understood as absolutely distinct. Let us take the case of knowledge. If we want to characterise knowledge, then we need to emphasise what is distinctive to knowledge. It would not do to pick out the common element between knowledge and error or mere opinion, the common element in this case being a belief. This will be clear from the fact that the explanatory role that knowing plays is distinct from the explanatory role that belief plays. And so it would not do to mention the common element that knowledge shares with error or mere opinion to characterise knowledge. The main task of naïve realism is to deny that appearances (which are the common elements) can have any 'autonomous status within one's mental economy'. Tim Williamson has tried to do this by showing that the common element thesis cannot be right. Williamson says that the very understanding of perceptual delusion is dependent upon recognising the element that delusions lack but veridical cases have. Under this thesis it is in terms of the veridical cases alone that one can make sense of the delusive case. Let us see how the argument runs. Under this view the veridical case is basic in the sense that it has a fundamental role in explaining what it is like for one to perceive something. And not only

---

119 In Mind, July 1995.
120 Mike Martin, 'the Reality of Appearance', p.9.
that, it also has a fundamental role in explaining what it is like for a person to merely have a hallucinatory or an illusory experience. We shall see how this is demonstrated. Hallucinatory and illusory cases, as we have noted, are indistinguishable from the veridical one. We have also noted that the conscious character of all these states (delusive and non-delusive) will seem to the subject to be the same if she were to reflect on it. In each case she will think that she is in a genuine situation of perceiving. So each of these cases to the subject is indistinguishable from a genuine case of perceiving, even the genuine one, though trivially. And this triviality is important for us. If we were to ask about the illusory and the hallucinatory cases, why do they seem to the subject like a veridical case, we will answer, because they are indistinguishable from the veridical case. But in the veridical case we do not need to appeal to the fact that this experience was indistinguishable from the veridical one in order to explain why it feels to the subject like a veridical case. There is, as though, no way of understanding how it is like to be in an illusory or hallucinatory state 'except by implicit reference to the kind of veridical perception from which it is indistinguishable'\(^{121}\). We can counter the charge that the internalists bring against the externalists saying that they cannot explain the cases of empty thoughts and behaviour, in the same way. We can say that there could be no understanding of these delusive cases without any reference to the non-delusive ones from which they are indistinguishable.

Martin says that this brings out clearly what the naïve realists (and in our case the externalists) need to deny. The naïve realist and the externalist do not actually need to deny that there is a common mental state invariant to the delusive and the non-delusive case. All that they

\(^{121}\) 'The Reality of Appearance', p.9.
have to point out is that this state has to be 'unavoidably' picked out and explained 'by reference to a non-common element, that is it will be explained by reference to some case of veridical perception'\textsuperscript{122}. We do not need to appeal to subjective indistinguishability to pick out and explain this state. This brings to doubt the subjective indistinguishability claim that we have aimed at questioning. But as we have said before, doubting this could get us into serious troubles regarding the question of first person authority or self knowledge and so let us now see how those problems could be solved by an externalist.

2.7 Externalism and Self-Knowledge

I have tried to argue in the last few sections against the incorrigibility thesis, the infallibility thesis and against the subjective indistinguishability principle. Rejection of these three theses would obviously involve a view that we could be mistaken about our own mental states. So in criticising internalism I have ended up in giving up one of the most intuitive ideas one could have: the idea of first person authority. This has been one of the greatest challenges that the externalist thesis faces and mine is no exception to it. Well, how do we get out of this messy situation? First of all, we have to find out if at all we are in a messy situation, i.e. if at all externalism leads to a rejection of self-knowledge and first person authority. And then see if an externalist can indeed get out of this mess. Let us first consider why it is believed that externalism cannot explain self-knowledge.

The reason for which externalism faces this challenge is not just that it is compelled to reject the infallibility or incorrigibility thesis. There

\textsuperscript{122} Ibid.
seems to be a sort of conflict between the immediacy and authority that are attached to self-knowledge and the broadness of externalistic content. It appears to many philosophers that under an externalist thesis we cannot make sense of the fact that we know our thoughts in a direct way. Does externalism really face this problem? 'If the contents of our attitudes characteristically depend on relations in which we stand to others in our speech community and to objects and kinds in the material world, then what states of mind we are in at any given point is going to be a function of prevailing empirical circumstances of a kind to which, in our ordinary effortless and groundless self-ascriptions, we pay absolutely no attention.' If contents of thoughts are indeed determined by its relational properties, then how can we explain the ‘effortless and groundless’ nature of our self-ascription? This is the central question here. Does externalism provide us with any compelling reasons for giving up our commonly held beliefs about self-ascriptions? If externalism is true, then it seems that we shall have to give up thinking that we know our minds in the way we think we do. So we see that having conceived of content in the broad way externalism faces this terrible problem. There are several lines that the externalists might take in order to meet the challenge brought against it on account of self-knowledge. It might say that the traditional treatment of self-knowledge claims are faulty, and if we can modify them then we can accommodate externalism with self-knowledge. It can also claim that the kind of authority we claim about our own mental states can be extended to external conditions of contents, and hence even if we take thoughts of ourselves to be externally determined, we can have authoritative access to them. An externalist can also come up with a denial of a tension

between externalism and self knowledge by saying that the kinds of claim that we make about self-knowledge are all wrong.

To start with, let us see what kind of claims we make about our knowledge of our own minds. It seems to be intuitively obvious that we have knowledge of our thought and this knowledge does not always depend upon any inference from other beliefs. This knowledge is characterised with a kind of authority and immediacy which is uncharacteristic of the thoughts that are conceived relationally. Given this intuition about self-knowledge it seems problematic to have self-knowledge if we were opting for an externalistic view of content. There may be two ways in which this common intuition about knowledge of our own thoughts could be understood. Either we know our thoughts on the basis of some inner observation or they are self-justified. It seems that with the externalist view about content it is implausible either way. On the one hand, as we are supposed to individuate the content properties of a thought by the relational properties of it, it is impossible to know what we are thinking merely by looking inwards. On the other hand, there are serious difficulties in taking our knowledge of our thoughts as based on nothing at all. We have to remember that the problem in accepting self-knowledge for externalists is not just a matter of accounting for the apparently unacceptable infallibility and/or incorrigibility theses. Even if the externalists manage to demonstrate that self-knowledge cannot be incorrigible, let alone infallible, still it will have one problem remaining for it to solve. It is this: 'sincere first person present-tense claims about thoughts ... have an authority no second- or third-person claim ... can have'. How can an externalist account for

---

See Donald Davidson, 'Knowing one's Own Mind', in Quassim Cassam (ed.), Self-Knowledge, Oxford University Press, Oxford, 1994, p. 44.
this unless it sees our knowledge of our own thoughts as distinct from our relational knowledge of thoughts regarding objects existing outside?

Traditionally exceptional claims are made about our capacities of knowing our own minds. We are thought to have a privileged access to our own mind rather than anything else in the world. But maybe this is only true of some of our thoughts, and this we have already seen. In the face of all this the notion of first person authority too seems to be dubitable. But surely in a very significant sense the way we know our own thoughts is very different from the way we know the thoughts of others. In the history of philosophy, we have seen that we were more concerned with making sense of our knowledge of other minds. And this had led us to behaviourism, which says that we could infer the state of other people's minds by observing their behaviour and by comparing them with our own mental states and behaviours. Now we are faced with an even greater problem: How can we know our own minds? We had previously thought that we could understand others as we believed that others thought like us. And we were more or less sure about what our own thoughts were like. But on the basis of what evidence? Definitely not the kind of evidence we take into consideration when we try to understand the thoughts of others. To take an example, if I see my friend's face gloomy and eyes filled with tears, I know that she is sad. But in order to know that I am sad I do not need to go in front of a mirror and see if I have those physical signs of sadness. So it seems that in case of other people's thoughts we infer, whereas in our own case we have no need for inference. We have as though a kind of direct access. Even in cases where we have inferential evidences in favour of our beliefs about our own mental states, we do not take help of them at all.
We shall first discuss the view which denies the possibility of authoritative self-knowledge altogether. Ryle\textsuperscript{125} tried to counter this common intuition about self-knowledge. Attributions of beliefs to ourselves are made in exactly the same way as it is done in the case of attribution of beliefs to others, according to Ryle. But this view is incapable of explaining many things. For instance, I entertain the thought that Quine's 'Two Dogmas of Empiricism' is one of the classic essays in philosophical literature. As soon as I think this, I am aware of what I think. How can Ryle explain this? It is not as though I deduced that I had that thought from some behaviour of mine, because it could well be the case that that thought had not led me to behave in any particular way as yet. So I came to know what thought I was entertaining non-inferentially.

An internalist conception of justification about our knowledge about our own mental states too tries to argue on this line. That is, under an internalist hypothesis our knowledge of our own thoughts cannot be inferential. We have a privileged access to them and if they seem to us to be the same then they are indeed the same. They say that the way we know our own mental states is absolutely different from the way we know about others. In the case of others we employ inferential methods to decide whether or not one is in a particular mental state, whereas in our case we know how it feels and that is all that there is to knowing about it. The intuition that the internalists have behind this thesis is that one can never be justified in holding a certain belief if it was the case that judged from the inside she could not be regarded as epistemologically rational or responsible for holding such a belief. The internalists provide us with a lot of examples in which the person perfectly satisfies the

\textsuperscript{125} In \textit{Concept of Mind}, Penguin Books, Middlesex, 1983.
external conditions of justification for a belief, but fails to have any internal justification at all. The internalists suggest that in order to be fully justified in believing something, the subject has to realise the connection between her belief and the evidence for her belief. This thesis of the internalists faces the regress problem of justification, just as any other theory of justification faces. If we justify one belief by another, then we have to justify the justifying belief first and this would lead to vicious infinite regress. The usual options that are open to such theories of justification are Foundationalism and Coherentism. But there is a special problem that an inferential conception of self knowledge faces. Boghossian says that this problem is independent of 'the standard problem of the regress of justification'\(^\text{126}\). Suppose that the belief \(p\) is justified relative to the belief \(q\). In the case of self-knowledge the belief \(p\) would be a belief about one's own belief, like the belief that I believe that \(r\). If we are to accept that all such beliefs are justified inferentially, then we have to say that it depends upon another belief, say \(s\). The subject has to know that she believes in \(s\) if \(s\) has to count as a justification of her belief that she believes that \(r\). Let us discuss this more systematically. If I am to be justified in believing that I believe that \(r\) then I shall have to satisfy the following conditions.

1. I believe that I believe that \(r\).
2. I believe that \(s\).
3. The proposition that \(s\) justifies the proposition that I believe that \(r\).
4. I know that I believe that \(s\).
5. I know that a belief that \(s\) justifies the belief that I believe that \(r\).
6. I believe that I believe that \(r\) as a result of knowledge expressed in 4 and 5.

\[^{126}\text{See his 'Content and Self Knowledge' in Philosophical Topics, Vol. XVII, No.1, 1989, pp. 9-10.}\]
The regress is patently clear. This perhaps suggests that the intuition that we commonly have regarding self-knowledge is right, i.e. that we know our minds/mental states directly without being mediated by any kind of inference from other beliefs. So can we know our minds/mental states through a kind of inner observation then? If we can, then the internalist intuition would be vindicated. But there are serious problems there too.

It seems to be extremely intuitively plausible to think that we have a kind of inner awareness by means of which we come to know about our own thoughts. In a very significant sense we do not identify our thoughts and come to know about them merely by looking inside ourselves. People have found this observational model of self-knowledge not acceptable.

Burge has tried to reconcile self-knowledge with externalism by coming up with a notion of 'basic cases' of self-knowledge. The first thing that Burge wants us to realise is that it is wrong to take self-knowledge to be a species of cognitive achievement. The distinctive feature of self-knowledge is that we are capable of thinking of those thoughts in the 'second-order verifying way'. What Burge is trying to say is that there are cases of judgements which are self-referential and self-verified. For instance, suppose I make the judgement I am thinking that I will have to learn how to make a Black Forest gateaux. If I make such a judgement, then I surely am having such a thought. These thoughts seem to be logically self-verifying. It is like if one were to utter the words 'I am talking', then his utterance would automatically be true in a sort of logically self-verifying way. Burge is suggesting that there are such self-regarding and self-verifying judgements which would count as the basic
self knowledge. Burge says that suppose an agent's thought to the effect that $p$ has certain constituents. We know that when one knows that one is thinking that $p$, one is also thinking that $p$. Consequently all the constituents of thought of $p$ are also carried over to her awareness or knowledge of her thought of $p$. This would be true even in the cases where the constituent conditions are externals. These constituents can never obstruct self-knowledge. If in the individuation of the thought of $p$ we need to make an essential mention of an external object, then we would have to make a mention of it also when we are individuating the knowledge of the thought of $p$. So, on a Burgian view, self-knowledge does not pose any problem for externalism in the first place.

Burge points out that the internalists have followed a Cartesian line regarding self-knowledge. Descartes definitely over-rated the capacity of the mind. We depend on observation of our own behaviour and reliance on the report of other people about us. As we have already discussed, Burge tries to say that not all the instances of self-knowledge that Descartes spoke of are really authoritative in the way Descartes wants and there are only a few which are and Burge calls them the basic cases of self-knowledge. And as we have already said, they are the self-verifying cases. If we are ready to accept the Burgian line, then we might have a solution to the problem of self knowledge which would not be inconsistent with the externalist thesis.

In a sense Burge's observation seems to be appealing, but how much of our self-knowledge would count as basic? It seems that at least in the cases of our more stable and well-entrenched mental states like beliefs and desires which we have been having for quite a length of time,

---

127 'Individualism and Self-Knowledge' in Cassam (ed.) *Self Knowledge.*
we can say that they do not count as basic self-knowledge. We might make judgements about them, but merely making judgements would not guarantee their truth. One thing can be said in favour of Burge and it is this: Though it is true that his thesis applies to basic cases of self-knowledge and yet this itself demonstrates that externalism poses no obstacle to self-knowledge in principle\textsuperscript{128}. Another thing that we should remember is that, under the Burgian thesis, one does not have to know all the ‘constitutive conditions that go into the thought that \( p \) being the thought it is in order to have that thought. And so if one is unaware of the various external factors that go into its being that thought, it neither follows that one doesn’t have that thought nor, given their common necessary conditions, that one doesn’t have the iterated thought that expresses self-knowledge of the thought\textsuperscript{129}. What Burge wants to point out is that the Putnam twin earth cases and the externalism that they lead to do not directly rule out the possibility of self-knowledge.

Donald Davidson\textsuperscript{130} has a different way of countering the apparent threat that externalism faces vis-à-vis self-knowledge. Davidson’s main line of argument is that we are mistaken about the source of the authoritativeness of self-knowledge, and that is why we think it to be incompatible with externalism. Once we realise this mistake, we shall see that in fact externalism can be well accommodated with self-knowledge.

To start with, we have to keep in mind that Davidson denies extreme attitudes regarding the question of contents of thoughts. He

\textsuperscript{128} See Akeel Bilgrami, ‘Can Externalism be Reconciled with Self-Knowledge’ in Pessin, Goldberg (eds.), The Twin Earth Chronicles, M.E. Sharpe, Armonk, New York, 1996.

\textsuperscript{129} Ibid., p. 367.

\textsuperscript{130} See his ‘Knowing One’s Own Mind’.
neither believes that following a Cartesian line we should take contents to be solely narrow, nor does he subscribe to a thorough going non-Cartesian thesis of the McDowell-Evans-strong-singular-thought-theory kind. He says that mental states are "inner", in the sense of being identical with states of the body, and so identifiable without reference to objects or events outside the body; they are at the same time "non-individualistic" in the sense that they can be, and usually are, identified in part by their causal relations to events and objects outside the subject whose states they are. Davidson claims that this kind of externalism is well in tune with self-knowledge. And that it follows from the fact that this thesis manages to show that first person authority can even be claimed about states which are partly identified by their relations to the objects outside the mind and the environment in general. And consequently even if our thoughts have wide content, they can be regarded to be equally knowable to us. The real argument that Davidson offers is that the authoritativeness of self-knowledge does not come from the kind of object that is revealed in it (inner or outer). The authoritativeness comes from elsewhere and if that condition is met, then thoughts with broad as well as narrow contents will be equally accessible to the subject. Now let us see how Davidson demonstrates this.

Davidson is aware of the fact that it is normal to assume that if a thought is identified in terms of its relation to external objects and environment, then it is unlikely that the subject could know such a state. Now why is it natural to assume this? The answer is simple: 'Because the external relation is not determined subjectively, the subject is not authoritative about that. A third person might well be in a better position

131Ibid., p. 48.
than the subject to know which object the subject is thinking about, hence be better placed to know which thought it was. It is true that people can find out what they think by finding out what their thoughts are about. For instance, once educated in chemistry Ed and Ted would be able to know what their thoughts were about when they were in a different world. But this kind of knowledge of one’s own thought is not characterised by the kind of authority that self-knowledge is regarded to be characterised by. And so the problem of self-knowledge looms large for the externalist. The problem is actually how to explain the obvious asymmetry between the way in which we know our own contemporary thoughts and the way that others know it. The only solution, one might think, that is open to us is the internalist one. If we made the content of thought narrow, then there would be no threat of loss of authority or immediacy. But is this the only alternative left for us?

Davidson says that there is a way open for the externalists. And this way depends heavily on the question: In which way do the externalists take the identification of mental states to depend upon external factors? If an externalist theory takes thoughts to be identified by objects (like propositions, representations) which are ‘before’ the mind, and if they take those objects to contain external objects and events, then of course about such thoughts one cannot claim authoritative self-knowledge. But if one rejects the antecedent, then one may not be forced to accept the consequent.

Davidson says that we should speak of a different way in which identification of mental states is done in terms of external objects. Burge

---

132 Woodfield, introduction to Thought and Object, p. 49
suggests one such way. Burge's arthritis example suggests that one can well have a propositional attitude with a content which has elements over which the subject does not have full mastery. Just as we might have only a partial knowledge of our own vocabulary, we might well have only a partial understanding of our propositional attitudes. It is wrong to suppose that we need to have a full understanding of the content of our thoughts in order to have those thoughts with those contents. Davidson agrees with Burge on this point and says that 'what we mean and think is determined by the linguistic habits of those around us'. This of course again brings first-person authority to doubt. But here again, Davidson claims that if we are careful enough in deciding how exactly does the linguistic habits of those around us determine our speech and thought, then we would not face the threat once more.

Let us see how Davidson thinks the linguistic practice of our fellows affect our speech and thinking. He says that a speaker (unless she has some perverse desire to perplex her audience) intends to be understood and hence intends to be interpreted in a particular way. For this she has to take into consideration the knowledge of the language of her audience. Davidson says 'It is the requirement of learnability, interpretability, that provides the irreducible social factor, and shows why someone can't mean something by his words that can't be correctly deciphered by another'. This makes a strong enough case for broad content but what remains a problem is that it is only to the narrow content that we can assign the Cartesian property of being known to the subject with authority and immediacy. This belief in the authority and immediacy of the awareness of the subject regarding the narrow content

---

133 See his 'Individualism and the Mental.
134 'Knowing One's Own Mind', p. 55.
of her thoughts is due to the following beliefs: 1. The subject can have a direct access only to what is in the head; 2. If thought is identified relationally (i.e. in terms of wide content), then they are not strictly in the head; and 3. Thoughts which are internalistically individuated are in the head. In other words, the content of the thought has to be in some way intrinsic so that the subject can have immediacy and authority with respect to it. This is a key internalist idea. And unfortunately some of the externalists too have held on to this idea and this has led them to face the problem of self-knowledge, among other problems. As Davidson says, it is only because it is commonly believed 'that ordinary propositional attitudes aren’t in the head’, that it is thought that ‘first person authority doesn’t apply to them’. It is this belief that Davidson wants to counter. He wants to show that it is wrong to suppose that if meanings and thoughts are identified partially by relations that we have with external objects and events, then those meanings and words cannot be in the head or in other words intrinsic to us. A simple example could demonstrate the absurdity that is involved in this belief. Sunburns are definitely intrinsic conditions of the skin, but they would not have existed unless the sun did. I cannot distinguish a sunburn from any other skin condition without referring to the sun which caused these burns. But this does not make the sunburn extrinsic to my skin. It is as much a part of my skin as my complexion is. All this proves is that it would be wrong to assign narrow content a privileged status of being intrinsic. And if

---

135 We shall be speaking about what could constitute the intrinsic aspect of a thought, and also if something external can constitute such an intrinsic aspect of a thought, time and again in this thesis. We do this especially in the following two chapters. What we find here will be crucial for the points I make in those two chapters.

136 ‘Knowing One’s Own Mind’, p. 58.

137 One might think that this view would discredit theories which speak of identities between the physical and the mental, but Davidson has tried to show how it does not. I shall not go into a discussion of this as it is not essential for our discussion here.
wide contents too could be intrinsic, then we could be equally authoritative about them.

Davidson says that there is a picture of mind which has been so instilled in us that it is difficult to get rid of this way of thinking about the mind. He says in this picture appearances are presented to the conscious self. These appearances are no part of the world, but they are representatives of the world. And the real problem is how to bridge the gap between the inside and the outside. In the case of the mental states it is the propositions or representations which are regarded as the intermediaries. And it is thought that the mind 'grasps' these. Davidson attacks this idea.

However we have to note that for Davidson when one entertains a thought, the authority does not come from the nature of the object that is 'before the mind' (along with its constituents). It comes from elsewhere. Davidson's main contention is that 'interpreting another person and attributing thoughts to her requires the assumption that she has direct and non-inferential access to her own thoughts'.

Of course there are philosophers who would not be happy with Davidson's response. I shall not go into a discussion of their views here. From what we have discussed so far it is clear that, as Burge and Davidson have both tried to show, it is not quite obvious why we could not have self-knowledge with respect to our relational thoughts. If there is any truth in this line of thinking, then externalism cannot be said to face the threat of not being able to account for self-knowledge.
2.8 Subjective indiscriminability as good evidence

From the discussions we have had in the last few sections it is clear that subjects cannot be said to have a special kind of privileged access to their mental states, as far as being unfailingly correct about them is concerned. I have tried to show that it would be wrong to assume the infallibility thesis and the incorrigibility thesis as right. And so we are equally vulnerable to mistakes about ourselves, as we are about others, and as we are about the outside world. In face of all this, we might now ask whether subjective indiscriminability can be regarded as a good evidence for the sameness of conscious mental states. As we have seen in earlier sections, the infallibility thesis which says that we are never wrong about our own mental states cannot be accepted. And this is partly because the infallibility thesis relies on a dubious notion of phenomenal characters. There is however a strong appeal of the subjective indiscriminability thesis. I think it is because as a methodological strategy it seems to be highly plausible. i.e. in most cases, if the subject cannot distinguish between one of her mental states and another, then the two states are type identical. This is a general methodological strategy taken in most cases where the question of sameness and difference arise. It is regarded as a good methodological principle that if two things are so similar that we cannot tell them apart, then they are similar unless proven to the contrary. This simple truth is borne out by our humdrum experiences. For instance, it is because we are usually right in detecting that there is water in front of us, that we get fooled by a mirage. Again it is because in most cases I can tell the difference between a snake and a rope, that an illusion of a snake in a rope, makes me run. All this is because indistinguishability is often a

\[\text{Akeel Bilgrami 'Can Externalism be Reconciled with Self Knowledge?', p. 366.}\]
good guide to the sameness of things. Seeming absence of difference is more often than not accompanied by real absence of difference. This has led us to take up a methodological principle: If we cannot see difference, then it is more likely than not that there is no difference. But this is as far as we can go. We should not make this methodological principle into a metaphysical principle, like: If we cannot see difference then there is no difference. The infallibility thesis was mistaken because it takes the general methodological principle beyond its methodological realm and makes a more full-blooded metaphysical claim about identity.

However the internalists can hold just on to the methodological claim and say that in most cases of mental states subjective indiscriminability is good evidence for the sameness of mental states. And because it is such a good evidence, internalists might think that while individuating mental states, it is how the state appears to the subject which should be counted as more important than anything else. But I think that a merely methodological claim cannot by itself (without any metaphysical or epistemological supports) help us to make claims about how an entity ought to be individuated. And so the internalists cannot just rely on their good methodological commonsense to determine their views in metaphysics. In order to make their methodological insight pay such rich dividends, the insight would need metaphysical or epistemological crutches. A stronger internalist thesis was just doing that, i.e. providing this methodological insight with metaphysical and epistemological foundations. This came in the form of Cartesianism and their belief in phenomenal character. But we have seen how faulty these two notions can be. And without their support this methodological insight (in spite of its plausible character) fails to get internalists too far
There are other problems that might plague this methodological thesis itself. The main worries with this methodological insight come from the numerous examples where how the state appears to the subject has failed to be a good guide to the nature of the state. The opponents (the internalists) need to provide an over-riding reason to say that the other cases (where they claim that how the state appears to the subject is indeed a good guide to the nature of the state) are not such cases as well.

2.9 What we can conclude

The above discussion shows that if (a), i.e. we can consider two mental states to be the same (i.e. type-identical) if the subject cannot discriminate them, can be a presupposition that might have motivated the internalists, then the internalists face the above problems. This is mainly due to the fact that the notions of identity and subjective indiscriminability in general (and in the particular case of mental states) do not coincide. In the next chapter I shall try to show that the other presupposition with which the internalists might be working, viz. (b), i.e. we can consider two mental states to be type-identical if they give rise to the same behaviour, ceteris paribus, is also a problematic presupposition.
Chapter Three
Behaviour and Externalist Explanation

In this Chapter I shall be discussing the second assumption that I believe that the internalists are operating with. The assumption is:

*We can consider two mental states to be type-identical if they give rise to the same behaviour, ceteris paribus.*

First of all I shall be showing how internalists indeed have this assumption and then I shall be trying to show what exactly it is to have such an assumption. Finally I shall say why I think that such an assumption might not be acceptable after all.

3.1 The Assumption

Let us start with a bit of background fact-finding. In fact, there is nothing wrong with the assumption as such. An externalist will have no problem with the assumption that the sameness of behaviour entails the sameness of mental states. Actually the problem that the externalists have is that they cannot accept the assumption as conceived by the internalists. This is due to the kind of notion of behaviour that the internalists employ in the assumption. Otherwise, the assumption actually embodies a common insight about mental states. To start with, we know that the need for having mental states in our story of the world was felt when we needed to explain behaviour. And so the central role that mental states played in explanation of behaviour was regarded as
one of the key features of a mental state. This is one of the reasons that we find a close link between mental states and behaviour. There is another reason for which behaviour came into the discussion of mental states. Earlier on, philosophers believed that we could know about mental states purely by introspection. But with the realisation that there were many serious problems with the introspective method, philosophers started to think that we could know about mental states by looking into the role that those mental states played in explaining behaviour. This close relation that mental states are taken to have with the behaviours that they explain, might have been the reason behind having this assumption in the first place. Philosophers also felt that the evaluation of the worth of a theory about the mind was tied with the question of how best could the role of the mental states in the explanation of behaviour be captured in that theory. In fact, we find philosophers taking into account the different roles that the mental states play in explanation of behaviour while classifying them into different kinds (e.g. beliefs, desires etc.). It is obvious that these classifications are closely tied with the question of individuation of mental states. Now psychological explanations of behaviours in terms of mental states actually presuppose causal interactions between mental states and behaviour and other mental states. This has made philosophers to think that we need to define mental states in terms of their causal powers. And it is here that the internalist externalist controversy comes in.

The question around which the controversy arises is whether to classify mental states like intentional states we solely have to look into what goes inside the subject, or do we have to, at least partly, look into what the environment is like. The internalists think that in order to define mental states in terms of their causal powers we have to
individuate them internalistically and not externalistically. And here the assumption has become different for the two contending parties. We find them differing even in their conception of behaviour, and consequently differing about what one should regard as the explanandum of psychology. We shall see that the reasons for their taking any particular stance about what could possibly be the explanandum of psychology are related to their conception of psychology itself. Internalists and externalists differ in their views about mental states, about behaviours (the explanandum of psychology) and psychology itself. All these notions are intertwined in such a way that one's views about any one of these notions have immediate bearing upon one's views about the others.

Let us now consider the assumption. As we have already said, our first task is to show that internalists do indeed have this assumption. To refresh our memory, the way that Segal and Fodor try to argue against Tyler Burge shows that they have this as an underlying assumption, i.e. that the sameness of behaviour provides good reason for the sameness of mental states. They say that since Ed and Ted behave in exactly the same way in earth and twin-earth in the presence of water and twater respectively (and also since they would behave in exactly the same way if they were to change places, unbeknownst to them), they must be in type-identical mental states. They go on to say that we could give a full explanation of behaviour in terms of mental states individuated purely internally and we do not need any essential mention of the objects towards which those mental states are directed.

The main problem with this assumption is that the notion of behaviour that these philosophers are working with is already a narrow
or an internalistic one. Consequently, an externalist can well say that it is because they have a narrow notion of behaviour, the internalists think that a purely internalist individuation of mental states is sufficient for the explanation of behaviours in psychology. The externalist would say that had the internalists employed a broad/wide notion of behaviour, then narrow mental states would prove to be insufficient for their explanations. In response to this externalist charge, the internalists have tried to demonstrate that it is this narrow notion of behaviour which we employ in psychology and it is this kind of behaviour that have to be explained and can be explained in psychology. And since an internalistic individuation of mental states rather than an externalist individuation is sufficient for the explanation of narrow behaviour, internalism is the right thesis about the mind and the mental.

The two main concerns for us in this chapter are to find out what is the notion of behaviour that we must adopt and what should be the nature of psychological explanation. And in knowing the answer to these questions we will be able to have a clearer view about how mental states are to be individuated — an issue which is at the very centre of our entire discussion in this thesis. As we have already noted in the very beginning, the two questions are, in fact, linked together. We shall see that the notion of behaviour that a philosopher opts for is largely influenced by his notion of psychological explanation or his belief about what is it that we are trying to explain in psychology and vice versa.

I have found arguments going both ways. To start with we can discuss Jerry Fodor\(^\text{139}\) and also Tim Crane\(^\text{140}\) who try to show that it is

---

impossible for us to reconcile causation with content, if we take psychological states to be wide. And since psychological explanations should be causal explanations, what we are explaining cannot be wide and consequently mental states should not be individuated widely, i.e. sameness of behaviour narrowly conceived indicates sameness of mental states. We shall also consider Jaegwon Kim’s thesis of psychophysical supervenience, which Kim thinks, has to be respected, and if it is respected then there is no way that we can have external mental states and externalistic explanations in psychology. What Kim tries to show is that externalist explanation is redundant. The supervenience thesis is at the centre of most of the discussions in this area and we shall see how Fodor too has emphasised its importance in the discussion of psychological explanations.

Externalists like Peacocke\textsuperscript{141} have brought forward arguments against the internalist by showing that psychological explanation can be externalistic. And if that is so, then the way that Fodor, Crane and Kim have tried to argue would not be acceptable. In addition to this we have externalists coming up with arguments in favour of a wide notion of behaviour\textsuperscript{142}. I shall be discussing how an internalist thesis might be compelled to have a very restricted notion of behaviour, which would make behaviour identical with bodily movements. I discuss Fred Dretske’s work here in order to show that bodily movements cannot be taken to be the explanandum of psychology. I shall be discussing Fodor’s argument against wide content/behaviour and I shall try to counter his contention following Peacocke’s argument in ‘Externalist Explanation’,

\textsuperscript{140} See his ‘All the Difference in the World’.
\textsuperscript{142} See Jennifer Hornsby’s ‘Physicalist Thinking and Conceptions of Behaviour’, in Pettit & McDowell (ed.), \textit{Subject Thought and Context}. 
and then go on to see if a wide notion of behaviour can be supported at all. I feel that Peacocke's argument might well apply against Kim.

But before going into this controversy we might just stop and consider a radically different view which questions both the internalist and the externalist positions. There is a radically different view about what psychological explanation should be like: One might think that it is wrong to think that psychological explanation is necessarily causal. And if that is so, then the kind of argument which is designed to show that we cannot reconcile causation with wide content will prove nothing. We shall first consider this position and see if it is tenable. This is so because if it is indeed tenable, then we need not be concerned at all by the internalist's arguments.

3.2 Non-Causal Explanations

There are philosophers, like Anscombe\textsuperscript{143}, Kenny\textsuperscript{144}, Melden\textsuperscript{145}, who think that explanation of behaviour is non-causal. What they say is that when we give an explanation of the behaviour of a person, what we do is give the reason(s) for which the person had behaved in that particular way. And they say that the relation between reason and action is not a causal one.

There are other philosophers, like Davidson, who defend the view that rationalising in this way is a kind of causal explanation. What exactly is the relation between a reason and an action? When we give reasons for our actions, we try to explain what reasons we had for doing that particular action. The reason helps us to see why that particular

\textsuperscript{143}In her \textit{Intention}, Blackwell, Oxford, 1959.
action appealed to the agent and what led the agent to perform the particular action at that particular situation.

Davidson says that whenever anyone does anything for a reason, we can characterise her with a kind of pro-attitude and a belief. For instance, if I see a person wearing a heavy sweater on a warm and sunny day, I will think she has a reason to wear such clothes. She possibly believes that she has fever and that she feels cold and that the sweater will keep her warm. Possibly both the pro-attitude and the belief lead the agent to do the action. Davidson says that when we try to rationalise a person’s action, what we are looking for is the primary reason that she had for doing the action. How can we characterise a primary reason? Davidson gives a preliminary definition of a primary reason, viz.:

‘C1. R is a primary reason why an agent performed the action A under the description d only if R consists of a pro-attitude of the agent toward actions with a certain property, and a belief of the agent that A, under the description d, has that property.’

Now let us consider Davidson’s example. Suppose I flip the switch and by doing so I turn on the light, illuminate the room, caution the prowler, make my neighbours think that I am home and so on. But why did I flip the switch? I could have done it to caution the prowler or I could have done it to let my neighbours know I am home, but let us suppose I didn’t. It seems that the primary reason is that I wanted to turn on the light and by doing so illuminate the room. But this doesn’t seem to have the kind of generality that is required of a primary reason. What Davidson wants us to realise is that the event, the existence of which makes ‘I turned on the light’ true is not the object of ‘I wanted to turn on

---

the light'. When I turned on the light I turned it on at a particular moment and in a particular way. But when I say that I want to turn on the light, then I do not always have a particular moment of time in mind in which I want to turn on the light or any particular way.

A primary reason consists of the pro-attitude as well as the belief related to it. For instance, if I want to turn on the light, then I will flip the switch only if I believe that by doing so I would be able to satisfy my want. But we really don’t need to mention the belief which accompanies the pro-attitude all the time, as it is quite easily understood.

In trying to trace the primary reason of an action, we are trying to fit the action into a general rational pattern of behaviour of an agent. We are trying to flesh out a character trait of a person. We are rationalising. We assume that others like us are rational and so they must have good reasons behind acting in a particular way. While assigning pro-attitudes and beliefs to anyone, we are actually interpreting the person so as to make her beliefs, attitudes, actions cohere.

Some philosophers think that the concept of cause that is applied elsewhere cannot be applied in the case of the relation between reasons and actions. Their alternative view is that the main thing that one is trying to do while giving reasons for a person's actions is simply giving a re-description of it, which makes sense of it.

Some philosophers have concluded two things from the above which Davidson thinks does not follow at all:

\[148\]

1. Causes are distinct from effects and since to explain one's actions by giving reasons is to re-describe it, reasons cannot be causes.

2. As reasons trace actions in a pattern, they explain them.

Davidson is questioning both. He says that we should not think that reasons are not causes just because they are re-descriptions of the actions in question. To start with, we have already seen that reasons are beliefs and pro-attitudes taken together. Consequently they cannot be the same as actions. Then it is also true that events are often re-described in terms of their causes. About the second, Davidson says that it alone cannot do the job of explanation. It is not at all clear how talking about the pattern and the context which contains both the action and the reason can tell us how the explanation story goes. The pattern is not enough as there may be many such different patterns and it is only what is behind the pattern that matters. This is so because we need to find out what is behind the pattern as that would help us in finding some kind of regularity in behaviour.

Davidson takes up Melden's example of a driver's raising his arm to signal. When a driver is approaching a turn, if he knows he ought to signal, and if he knows that to signal he needs to raise his arms and knowing all this in that situation, if he raises his arm, then he is making a signal. Davidson says that all that this shows is that he has a reason to raise his arm. But was it this reason which made him actually raise his arm? Unless it is the intention to signal which caused the driver to raise his arm, the description 'signalling', Davidson says, would not manage to explain the action of hand raising. Davidson refuses to accept a 'mysterious connection' between reasons and actions. One may have a reason to act in a way and might act in that way for that reason. But
another person or the same person could have the same reason to act in that way and act in that way and yet not for that reason. And there is no way to figure out if one has done something for a particular reason unless the relation between the reason and the action is a causal one. To get a better example, we might take the example\textsuperscript{147} that Tim Crane gives in the Second Chapter of his book \textit{The Mechanical Mind}. Let us suppose that there is this man who desires to kill his brother and has a reason for it, viz. the brother is an impediment in his career advance, or in his inheritance. Suppose further that he goes to a bar and gets involved in a brawl and without knowing kills a man who is in fact his brother (may be the brother was in disguise). Now in this case can we say that the person killed his brother for the reason that he was in the way of his career advancement or inheritance? Obviously we cannot say that. Imagine an alternative situation where he knew that his brother would be in the bar and he would be in disguise and the killer plans everything meticulously and makes it so that it seems that he killed his own brother inadvertently. In this case surely we can say that he killed his brother for that particular reason. Now why do we say that? It is because in the first case the killer would have killed the brother even if he didn't have that reason. But in the second case it is precisely because he has that reason that he kills his brother, i.e. had he not had that reason he would not have killed his brother. And this definitely shows that a reason has got to be a cause. This is so because in the absence of the reason (all things being equal) there would not be the action at all. This is exactly how cause is seen: \( A \) is a cause of \( B \) if the absence of \( A \) (all other things being equal) entails the absence of \( B \).

Considering all this, Davidson concludes:

C2. The primary reason of an action is its cause.

Davidson actually has no knockdown argument, but we might find the causal picture more plausible for the above mentioned reasons and also because we want to place the mind in the natural world of causes and effects and look for causal explanations in the realm of the mental\textsuperscript{48}.

3.3 Internalistic Explanation

Now we have managed to show that the kind of psychological explanations that are sought and given are more often than not in terms of the causal interactions that mental states have with behaviours and other mental states. These causal explanations can be of different sorts. And the test of the merit of a theory of mind will rest upon its (causal) explanatory power. In fact this has led philosophers to define mental states in terms of their causal interactions. And so the role of mental states in causal interactions has played an important role in the individuation of mental states. Again, what the nature of psychological explanation will be, depends upon the way that mental states are individuated. To quote Rosenthal from his General Introduction to \textit{The Nature of Mind}\textsuperscript{49}, it 'is crucial to psychological explanation that we know when two mental states are of the same type'.

\textsuperscript{48} As Crane says in \textit{Mechanical Mind} (pg. 62), we could be interested in having 'mechanical/causal picture of the world' and there is no reason why throughout 'nature we find causation' and yet the mind would be exempted from this kind of picture.

We have already spoken (in the First Chapter) of a possible argument in favour of internalism. It is Tim Crane’s argument which says that if anything could possibly cause behaviour, it has to do so only by mediation of intrinsic properties of the subject. And since as far as the intrinsic properties of the two twins across worlds are concerned, there is no difference between them. Consequently, the external circumstances could not have caused the twins to behave in the way they do without affecting their intrinsic properties. And so when we speak of behaviour, we do not speak of wide behaviour but we speak of narrow behaviour, as this is the behaviour about which we can make such causal generalisations. This behaviour could be caused by mental states which are internalistically conceived. We will not discuss Tim Crane’s position here in detail. We shall give an argument against his thesis that if anything could possibly cause behaviour, it has to do so only by mediation of intrinsic properties of the subject. This we will do in the fourth chapter by saying that even what is external can be regarded as intrinsic.

3.3.1 Jerry Fodor’s Arguments

One main internalist argument against the externalists is that psychological explanation is all about picking out causal powers and there is no way of doing it if content is conceived widely/broadly and not narrowly. This is because the causal powers of mental states are not dependent upon the relation that the mental states bear with the external world, but rather the causal powers supervene on ‘local microstructures’ or internal properties. This is one of Jerry Fodor’s arguments for narrow

150 We find it in 'All the Difference in the World'. All this will be discussed again in detail in the following chapter.

151 We shall be discussing this line of argument once again in the Fourth Chapter. For the time being we shall concentrate on Fodor’s argument.
content. I would like to pick up this classic internalistic argument, as we find it in *Psychosemantics* and 'Modal Argument for Narrow Content'. Fodor starts with pointing out that if mental states of a subject are conceived externalistically, as it is done by common sense, then they would not supervene on the brain states of the subject. The thought experiments of Putnam and Burge demonstrate this. The mental states would supervene upon the brain states, if and only if every change in the mental states would be accompanied by a corresponding change in the brain states. But in the cases spoken of by Putnam and Burge, Fodor thinks, that is not what is happening. For instance, the twins in Putnam's example vary in their mental states (externalistically conceived, as their contents vary), yet they are in exactly the same brain states. This is something that Fodor finds unacceptable, as he thinks that it would not be possible for us to make a science of the mind if we did not 'respect' this supervenience.

We have to keep in mind that the violation of supervenience results from having two different standards of individuation for mental states and brain states. On the one hand, mental states are individuated relationally, on the other hand, brain states are individuated non-relationally. We could handle this threat by taking two radical stands. We could either say that both common-sense psychology and scientific psychology are externalistic and hence both view mental states and brain states relationally. Or we could take another radical stand and say that both common-sense psychology and scientific psychology are internalistic and that both take mental states and brain states non-relationally. Fodor doesn't take any of these radical stands. He says that there is a marked

---

152 This is an instance of how the notion of mental states that one adopts determines the kind of psychological that one will have; and it also shows how the kind of
difference in the way that common-sense sees mental states and the way that scientific psychology sees them. If brain states, like mental states, were individuated relationally (for instance, if we were to take a brain state as a state of the brain which is in a body which is in a world with distinctive things in it), then supervenience would have been retained. But people find it extremely difficult to think that brain states have to be individuated relationally. Fodor says that an externalist like Burge might see no problem in the violation of supervenience. One can say that both common sense and scientific psychology view brain states as non-relationally individuated and mental states as relationally individuated. And one can say that both common sense and scientific psychology violate supervenience and there is nothing wrong about it. But Fodor points out that there is something wrong with this picture. The main reason for finding this picture faulty is that the reasons for which one takes brain states as non-relationally individuated and individuable are the reasons which are there even for the mental states. In other words, one ought to take mental states as non-relationally individuated for the very same reasons as the ones that make them individuate brain states non-relationally. Now what are these reasons?

Fodor first wants us to look into the reasons behind our belief that the brain states have to be non-relationally individuated. Scientific theories are usually taken to be internalistic. This idea is closely related to the idea of causal power. In scientific psychology states are individuated in terms of their causal powers and it is only when states differ in their causal powers that they are taken to be different. What Fodor wants to demonstrate is that being in a world with H₂O or being in a world with XYZ does not make any difference in the causal roles of the mental states...
of the subject. And hence these two mental states should be regarded as identical rather than not. This is a clear indication that the mental states should be individuated non-relationally rather than relationally. In order to make his point he construes a case similar to the above one. It goes like this: Suppose you have a dime with two sides: a head and a tail. Let us define a property 'is an H-particle at t' in a way such that an object satisfies this property at t if and only if your dime shows its head face up. We define the property 'is a T-particle at t' along similar lines. Fodor says that you can place the dime with its head face up and you would thereby be making all the objects of the world satisfying the property of being an H-particle at t. Again, you could place the dime with the tail face up and thereby make all the objects in the world satisfy the property of being a T-particle at t'. By doing this one can also change all brains in the world into H-brains and T-brains respectively. What these cases demonstrate is that whether an object is an H-particle at t or a T-particle at t' will not affect its causal power. This is totally irrelevant to its causal power. And as in sciences we are interested in giving explanations in terms of the causal powers of things, there would be no place of this classification of objects into H-particle at t and T-particle at t' in such a science. As Fodor puts it, 'if an event e is caused by H-particle p, then the same event e is also caused by p' in the nearest nomologically possible world in which p is T rather than H'. And so Fodor concludes that the properties of being H-particle at t and being T-particle at t are of no use in classifying things for the purpose of scientific explanations. Fodor carries this over to brain states too. He says that in the same way it becomes evident that being in an H-brain state or being in a T-brain state makes no difference in causal powers. And then Fodor would carry this over to mental states. Being an H-mental state or being in a T-mental state would similarly make no difference in causal powers. And so Fodor urges us to understand that if
we are looking for causal explanations, then it would be useless to conceive mental states, or brain states as relationally individuated.

The main point of Fodor's argument is that the causal roles of $H$-mental states and $T$-mental states (in the Putnam example $H_2O$-thoughts and $XYZ$-thoughts) are the same. What exactly does this mean? Roughly speaking having same causal powers means having same causal consequences. One may not accept that $H_2O$-thoughts and $XYZ$-thoughts have the same causal power, as one may not accept that $H_2O$-thoughts and $XYZ$-thoughts have same causal consequences. Fodor is aware of this possible counter-attack and he tries to answer it.

Fodor considers the charge that Ed's utterance 'Bring water' and Ted's utterance 'Bring water' might have distinct effects. When Ed says 'Bring water', he gets $H_2O$, while when Ted says 'Bring water', he gets $XYZ$. So the utterances and the mental states that are behind them have distinct effects and consequently they differ in their causal roles. In answer to this, Fodor says that the evaluation of causal powers can never be made within a context, it has to be made across contexts. What does this mean? It means that if Ed was in Ted's context then his utterance would have the same effect as Ted's in that context and *mutatis mutandis*, if Ted was in Ed's context then his utterance would have the same effect as Ed's in that context.

But maybe we could come up with a better example. Suppose I am a chemist and I am also aware of the fact the Ed and Ted are from earth and twin-earth which have water and twater respectively. Further suppose, we are all in earth and Ed asks me for water and I bring him a glass of $H_2O$. And then Ted asks me for water and I bring him a glass of
XYZ (maybe from the tank full of XYZ that I had managed to bring from twin-earth while visiting it on a scientific mission). Surely 'Bring me a glass of water' uttered by Ed and 'Bring me a glass of water' uttered by Ted has two distinct effects: (1) my H₂O bringing behaviour and (2) my XYZ bringing behaviour, respectively. And as is required, my behaviour would change across the contexts. If I am in twin-earth and Ted says 'Bring me water', I will bring it from the tank of XYZ. But suppose Ed asks me 'Bring me water' in twin-earth, then I could well say to him 'Sorry we don't have that stuff around here' (supposing that I hadn't brought to twin-earth a tank full of H₂O). This demonstrates that the mental states associated with Ed's utterance 'Bring me a glass of water' and the mental states associated with Ted's utterance 'Bring me a glass of water' too have distinct effects. And consequently they differ in their causal power.

Fodor has also considered why externalists have problems in conceiving mental states internalistically. The main difficulty that externalist philosophers had with the notion of internalist content is that such a notion violated one of their basic beliefs: that contents of mental states determine extension. Externalists usually believe that a variation in the extensions is a sure indication of variation in the contents. It is because of this that externalists wanted to individuate the content of mental states externalistically. The twin earth thought experiments, the externalists thought, demonstrated that unless we took contents externalistically they would fail to determine extension. Fodor counters this externalist view. He shows that it is wrong to think that the twin-earth thought experiments demonstrate that contents need to be externalistically individuated in order that they manage to determine extensions. All that the thought experiments demonstrate is that contents (even internalistically conceived) determine their extensions only relativised to a
context. All that these thought experiments have shown is that contents were like functions which map contexts onto truth values. Fodor says that it is the relation that is there between Ed and earth (or between Ted and twin earth) which fixes Ed's water thoughts to H$_2$O (and Ted's water thoughts to XYZ). Fodor calls the condition that is satisfied by Ted and twin earth which fixes the context of Ted's water utterances and water thoughts as C. On similar lines, he calls the condition that Ed and earth satisfies which fixes the context for Ed's water thoughts and water utterances, C'. Fodor says that all that he has said demonstrates that if an organism shares all its neurophysiological structure with Ted and satisfies C then both his and Ted's thoughts would have the same truth conditions. And so Ed, who does indeed share with Ted the same neurophysiological structure, will have thoughts that have the same truth conditions as the thoughts of Ted, if he satisfies the condition C. Fodor adds that all that the twin earth stories show is that since Ted and Ed's contexts differ 'de facto', their thoughts have different extensions.

Although Fodor's thesis is extremely powerful, I feel there may be some externalistic objections to it. To start with we have to think if at all Ted can satisfy the condition C', and if at all Ed can satisfy the condition C? I think not. Let us see why I say this. What are the conditions C and C'? Fodor says that in order to articulate these conditions we have to bring in something about 'the constraints on the causal relations between H$_2$O tokenings and water-thought tokenings (and between XYZ tokenings and 'water'-thought tokenings)'. I would like to understand what Fodor wants to say by this. What kind of causal relation there must be between Ed's H$_2$O tokenings and water-thought tokenings so that when Ed has a water-thought it gets fixed to H$_2$O rather than XYZ? Could it be that Ed's water-thoughts are caused by Ed's interactions with H$_2$O rather than XYZ? And
so in a way Ed’s water thoughts are causally linked with \( \text{H}_2\text{O} \) whereas Ted’s are not and cannot be so linked, unless Ted’s ‘water’-thoughts resulted from his interaction with \( \text{H}_2\text{O} \). Consequently we can speak of Ted’s water-thoughts referring to \( \text{H}_2\text{O} \) in Ed’s context, only if Ted’s water thoughts were caused by \( \text{H}_2\text{O} \) rather than \( \text{XYZ} \). That is, it is the causal relation with \( \text{H}_2\text{O} \) rather than \( \text{XYZ} \) which goes into the generation of Ted’s thought. This definitely is an externalist conception of thoughts. It seems to me that Fodor will not be able to articulate these conditions \( C \) and \( C' \), unless he already subscribes to an externalist conception of mental states. Fodor might come forward with a counter argument at this point. He could say that my criticism was a sort of external criticism, and so it did not affect him. All that he wanted to demonstrate is that the difference between \( C \) and \( C' \) was only a result of the difference in the input (in one case \( \text{H}_2\text{O} \) and in the other \( \text{XYZ} \)) and output (in one case \( \text{H}_2\text{O} \) involving behaviour and in the other case \( \text{XYZ} \) involving behaviour). The brain states of the twins being the same, the mental states which supervened on them would remain the same in spite of these input, output differences. At this point we can say that we have reached a stalemate. Let us try to find other internal criticisms against Fodor’s thesis. But before doing so we have to consider his similar argument in ‘A Modal Argument for Narrow Content’.

In ‘A Modal Argument for Narrow Content’, Fodor is considering, once again, the question what kind of difference in water-thoughts and water-thoughts of the twins would count as a difference in causal power. He follows a Humean line in establishing his thesis here. He says that suppose we have a pair of causes \( C_1 \) and \( C_2 \) and a pair of respective effects \( E_1 \) and \( E_2 \). \( C_1 \) differs from \( C_2 \) as \( C_1 \) has the cause property \( CP_1 \) and \( C_2 \) has \( CP_2 \). \( E_1 \) and \( E_2 \) differ from each other because they have
different effect properties $EP_1$ and $EP_2$ respectively. The difference in $C_1$ and $C_2$ is responsible for the difference in $E_1$ and $E_2$ respectively. In other words, if $C_1$ had the cause property $CP_2$, then it would have given rise to $E_2$ with the effect property $EP_2$, and similarly in the case of $C_2$'s having $CP_1$. Now Fodor says that what is in question is this: which are the instances in which 'the difference between having $CP_1$ and having $CP_2$ is a difference in causal power in virtue of its responsibility for the difference between $E_1$ and $E_2$'?\(^{153}\) Fodor says that if we manage to answer this general question, then we would manage to decide whether the 'difference between having water thoughts and having twater thoughts is a difference in causal power in virtue of its being responsible for the difference between my producing water behaviours and my twin's producing twater behaviours'?\(^{154}\). Fodor comes up with a general principle which has to be satisfied by two causes in order that they are endowed with two distinct causal powers. The principle is: 'Only when it is *not a conceptual truth* that causes that which differ in that one has $CP_1$ where the other has $CP_2$ have effects that differ in that one has $EP_1$ where the other has $EP_2$'.\(^{155}\) Fodor tries to demonstrate that the difference between the water-thoughts of Ed and the twater-thoughts of Ted are conceptual rather than contingent. And hence they do not mark any difference in causal power.

Fodor, in his characteristic style, comes up with examples of conceptual and non-conceptual differences. We know that our sons can be nephews only if we have siblings. This is a necessary truth, as it is conceptually impossible for some one being a nephew unless at least one of his parents has a sibling. Suppose that all nephews have a contingent

\(^{153}\)In *A Modal Argument for a Narrow Content*, p.9.
\(^{154}\)'A Modal Argument for a Narrow Content', pp. 9,10.
property, say G. Now it is a necessary truth that if I have siblings, then my sons will have G, if all nephews have G. And it is a necessary truth in spite of the fact that nephews might have G only contingently. Fodor wants us to compare this case with a case which he considers to be completely contingent. Suppose I am a planet and contingently I have a Keplerian orbit. Suppose further that E is a property that all Keplerian orbits have contingently. Consequently if 'E is a property that Keplerian orbits have, then if I am a planet then my orbit has E.' This is a contingent truth as 'it is contingent that planets have Keplerian orbits, so it is contingent that if I am a planet then my orbit has whatever properties that Keplerian orbits have'.

Now let us consider the case of water and twater thoughts. Fodor says, right at the beginning, that on an externalist thesis I have a water thought only if I am related to water in the right way. This he says is a conceptual truth. And hence if B is a property that water-thoughts have, then necessarily my thoughts would have B provided I am related to water in the right way. And hence if we conceive water-thoughts in this externalist way, then we would not be able to say that the distinction between water and twater thoughts mark a distinction in the causal powers of these thoughts.

It is in a sense an extremely powerful argument. This is particularly so because externalists do have a tendency of saying that a water thought is almost definitionally a thought which has a relation with water or rather H2O. This is because they take the relation with H2O rather than XYZ as constitutive of the very identity of water thoughts.

159 'A Modal Argument from Narrow Content', p. 19. Italics mine.
156 'A Modal Argument from Narrow Content', p. 23.
And that is the reason why Fodor felt that this would only indicate that the difference between water thought and twater thought was a conceptual one rather than a contingent one (as one was definitionally linked with H$_2$O and the other was definitionally linked with XYZ).

Christopher Peacocke however has tried to demonstrate that there is something wrong in Fodor's argument. We find his argument against Fodor in his 'Externalist Explanation'\textsuperscript{157}. Peacocke wants to point out that there is a significant difference between the relational properties of a behaviour and the intentions that produce that behaviour. Peacocke has a brilliant reason for this. He says that it is not a matter of conceptual necessity that people who have water thoughts produce behaviours which involve relations with water. Let us go back to the twins. If Ed faced a bottle of twater in earth, he would desire it if he had a mental state like water-desire and this would result in twater related behaviour in him. Merely being caused by a water thought does not make this behaviour of Ed's a water behaviour.

We can also take Fodor's argument seriously and say that it is indeed the case that it is a matter of conceptual necessity that water thoughts were thoughts which were constitutively linked with water and involved water-related behaviour. Peacocke says that even in such a case it did not follow that this did not record a serious causal connection and that being related to H$_2$O rather than XYZ would not count as a causal power. Peacocke says that let us grant that the following conditional was \textit{a priori}:

\begin{footnotesize}\begin{itemize}
\end{itemize}\end{footnotesize}
'If the subject is in his normal environment which includes water thoughts, he will (in the presence of other suitable attitudes) behave in various ways in relation to water.'\textsuperscript{158}

This does not imply that being related to H\textsubscript{2}O rather than XYZ could not be regarded as a difference in causal power. Peacocke cites a similar case to make his point. He says, suppose that we have the following conditional which was, like the previous one, \textit{a priori}:

'If an organism develops normally, and one of its parents has a dominant gene for blue eyes, then it will have blue eyes.'\textsuperscript{159}

What is important is that in spite of the \textit{a priori} character of this conditional, it cannot be denied that facts about genes do get mentioned in the causal explanation of someone's eyes being of a particular colour. The fact that the conditional is \textit{a priori} does not entail that as an \textit{a priori} matter eye colour has got be explained genetically. On similar lines, we could say that despite the \textit{a priori} nature of the first conditional, it is not \textit{a priori} that water thoughts get mentioned in explanation of water related behaviour. If Ted was to have water related behaviour because he was unaware that he was switched to an environment without XYZ and still drank what he thought to be water, then in explaining his water behaviour we would not cite his water thoughts, we would cite his water thoughts.

3.3.2 Jaegwon Kim's Argument

What we find in Kim is an argument that shows that externalistic explanation is redundant. In his 'Psychophysical Supervenience'\textsuperscript{160}, Kim takes up one by one all the instances where one might think that an

\textsuperscript{158} 'Externalist Explanation' p. 211.

\textsuperscript{159} Ibid.

\textsuperscript{160} Ibid.
externalist individuation of mental states has to be invoked and consequently an externalistic explanation can be given. He shows in all those cases an internalist individuation of mental state and an internalist explanation is sufficient. The thesis that holds a key to the success of Kim's argument is his thesis of psychophysical supervenience. We have already seen the importance of this thesis in Fodor. Let us now see how Kim uses this thesis in the benefit of an internalist explanation.

First of all let us see how Kim characterises supervenience\textsuperscript{161}. In a very simplistic way we could say that the supervenience thesis says that if a person is physically exactly like you, then she will share your psychological life with you. This shows that ‘All psychological states and processes supervene on the contemporaneous physical states of the organism’\textsuperscript{162}. An externalist thesis might counter this by saying that I might have properties which I do not share with my physical replica and such properties might go into the determination of some of my psychological properties. This brings into doubt the position that all our psychological states can be said to supervene on our physical states. Instances of such states which are determined not by our internal properties but properties of our external environment are discussed by Kim. He considers–if my replica can ‘remember being strafed by a jet fighter in a war over a twenty years ago’; if my replica can think of Vienna in the way I do having gone there a couple of years ago; if my replica can feel glad in the way I do when I am invited to the Dean’s party and feel in the way I feel embarrassed that I could not remember the Dean’s first name. He considers if one’s replica who has been so


\textsuperscript{161} Kim has various formulations of the supervenience thesis. But for our present purpose it will be enough to take the thesis as it is described in ‘Psychophysical Supervenience’.
conditioned that it has a tree-seeing sensation in the brain at the time that the original sees a tree. In such a situation can the physically identical replica said to be seeing the tree in the way that the original person is. Kim finally considers actions that are performed by a person which involve complex networks of social practices and customs. It might seem impossible that the replica would be performing the same actions that the original does. These cases seem to demonstrate that not all of our psychological states can be said to be supervening on our contemporaneous physical (internal) states.

There can be two ways of handling this problem for the supervenience thesis. One can say that psychophysical supervenience can be claimed about only a subset of mental states. But this would not be an acceptable alternative for the internalist. The internalist can say something else. As Kim points out, an internalist 'can broaden the supervenience base ... to include a person's past physical history ... to include physical states of objects outside the organism'\textsuperscript{163}. Kim thinks that one has to bring in possible worlds in formulating the supervenience thesis in order to bring in these exceptional cases of mental states within the purview of psychophysical supervenience. And hence the supervenience thesis would then be that if two possible worlds are indistinguishable physically, then they would be indistinguishable in their psychological respect as well.

After this Kim goes on to show that the putative non-internal mental states will serve no purpose of psychological explanation which the strictly internal states do not. This is known as Kim's Explanatory

\textsuperscript{162} Ibid. p. 178.
\textsuperscript{163} Ibid. p. 187.
Thesis. What Kim tries to say in this thesis is 'Internal psychological states are the only psychological states that psychological theory needs to invoke in explaining human behaviour — the only states needed for psychology'. As we have already seen, Kim tries to demonstrate the plausibility of this thesis by considering the cases where we think that we need to invoke external objects in order to identify mental states and to explain behaviour that are generated by them. He shows that even in those cases only the internal states are enough for explanation. Kim says that we need not make any reference to non-internal psychological states in a folk-psychological theory. He says 'the causal explanatory role of any noninternal, psychological state can be filled by some internal psychological state'. And he says that if this can be proved in general, then that would definitely show that we need no non-internal psychological states for explanations in psychology.

Let us take Kim's example of a person's turning the knob of the burner of a stove. Now it is indeed sufficient to give an explanation of the act by bringing in the agent's belief that turning of the knob would lead to the turning of the burner. But Kim says that it is not necessary to have such an explanation. What causal role does my belief, to the effect that turning of the knob would result in switching on the burner, have in that case? Do any of the non-internal states that are mentioned in the explanation of a behaviour have a causal-explanatory role? Kim argues that they do not. What Kim is trying to say is: psychological states cannot serve to explain why one turned on the burner, because such states would never be able to explain why the burner went on. Moreover, for Kim, why the burner went on is not a matter we need to

--

164 Ibid. p. 183.
165 Kim's 'Psychophysical Supervenience', pp. 181-182.
psychologically explain. Rather, why one turned the knob is what needs to be psychologically explained. And the reason that Kim says this is: if one was to say that in this case not only did I turn on the knob but also the burner, then we would also have to bring in all the other things that one manages to do by turning the knob, like heating the milk, making coffee, serving coffee to my friend displeasing her by putting too much sugar in it and so on. We are instead in search of a primary cause or reason which had lead to the behaviour. So Kim is saying that the job of psychological explanation is done, when one manages to find what would explain bodily movements. In a simple way we might say that what Kim is appealing to is this: what is most contiguous to the effect is most likely to be the cause in this case. This, however, might not be acceptable to all.

We shall now be talking about Peacocke's views as we find them in his 'Externalist Explanation' where he also tries to prove how externalist explanation is possible and in fact is required in some cases.

3.4 Non-Narrow Explanations

We have another good argument from Peacocke. In this argument he tries to demonstrate that, contrary to the internalist's contention, we can have externalistic explanations of behaviour. And not only that, sometimes it is only the externalist explanation which is plausible. Peacocke in his 'Externalist Explanation' starts by saying that psychological explanation is often directed towards an explanation of the

---

Kim, 'Psychophysical Supervenience' pg., 188.
relational properties of an event. Suppose, we bring in mental states in giving an explanation of a behaviour of a person. Often such a behaviour is identified relationally. To take Peacocke's example, a hand movement may have different relational properties. For instance, it could be a movement towards the window or towards a person in the garden beyond the window, it could be a movement in the northward direction. All these are relational properties of this behaviour of hand movement. We will see later, when we discuss about behaviour and bodily movement, that it is not merely why the hand moved that we want to explain. We want to explain why someone moved her hand. And quite often what we want to know is why did a person move her hand towards the person in the garden or towards the window or northwards, and so on.

Peacocke reminds us that the fact that many philosophers took the explanandum of psychological explanations to be external or relational was 'a consequence of broadly interpretationist approaches to intentional content'\(^\text{167}\) (as is found in Davidson in his *Inquiries into Truth and Interpretation*). Peacocke explores what he calls a distinctive feature of explanations of any kind of relational properties. 'It concerns the nature of the counterfactuals normally sustained by these explanations'\(^\text{168}\).

Suppose, we are trying to give a psychological explanation of Jane's pointing towards Andy in the garden. The kind of psychological explanation that we can give is that Jane pointed towards Andy because she wanted to point at Andy. And because she believed that he is in the garden (possibly because she could see him) and because she believes that she can point her finger in his direction and make an ostention to

\(^{167}\) Ibid., p. 204.
\(^{168}\) Peacocke, 'Externalist Explanation', p. 205.
him, she pointed towards Andy. What Peacocke is saying in explaining her act is that, all other things remaining the same, had Andy not been in that position in the garden, Jane would not have pointed her hand towards that direction. And, as we see, that the formulation of the antecedent is made in environmental terms, and also the specification of the consequent involves the agent's relation with the things in her environment. In any case, we should keep in mind that we can have alternative relational explanations of the same event. Peacocke suggests a general thesis about externalist states:

'It is partially constitutive of the identity of any externalist state that, in suitable circumstances, it can explain, or be explained by, relational properties of external objects or events.'\(^\text{169}\)

The problem that internalists have with this kind of externalist notion of psychological explanation of external/relational explanandum in terms of an external/relational explanans, is that this violates the principle of psychophysical supervenience. Internalists believe that brain states are sufficient for explaining behaviour. And as externalistically conceived mental states do not supervene on brain states and as a result they would turn out to be epiphenomenal.\(^\text{170}\) What Peacocke points out is that this internalist worry is unfounded, because what is being explained in psychological explanation is not just bodily movement and so brain states would not be enough to explain them. And this externalist thesis would not violate the psychophysical supervenience either, as they were not concerned with brain states and the narrow behaviours that they cause. This shows how externalist explanation is possible, honouring

\(^{169}\) Ibid, p.206

\(^{170}\) We shall discuss this position in detail in the Fourth Chapter.
supervenience between brain states and narrow behaviour. He has a stronger argument at hand which comes out while he considers a possible counter-argument to his position just described.

Peacocke has considered a possible counter-argument from the internalists. An internalist might say that one could supplement internal states of the subject with some features of the environment that the subject belongs to. Such supplementing would ensure that the internal states along with those supplements would manage to explain the external properties of the subject's behaviours just as much. But the question is, what kinds of things do the internalist propose to supplement the internal states with? Suppose we are trying to explain why Jane was pointing at Andy. Peacocke considers some such possible supplement to the internal states that Jane has when she is pointing at Andy in the garden. The internalist might regard as the supplement, the truth that Andy is in the direction of Jane's hand movement. Peacocke will say that such a supplement will not be enough. This is so because there are other things which are in the direction of Jane's hand movement. For instance, the BT Tower might be in the direction of Jane's hand movement, but the psychological explanation doesn't have to be explaining that Jane is pointing towards the BT Tower. As Peacocke says, the 'explanation of a truth by a given set of states is not preserved by substitution of coextensive predicates in that truth, not even if we add a statement of the coextensiveness to the original explanation.' He adds that this general thesis is demonstrated by the fact that in the Jane and Andy case, if we supplement the internalist explanation, even then that would not sustain the counterfactual that if Andy was in a different direction then Jane would still point at him. Moreover, if the internalists go on adding supplements to ensure that they sustain those counterfactuals, then there
would always be a risk that the internal states would become actually external states.

Peacocke presents the internalist position in form of a dilemma. This dilemma emerges from the internalist claim that if an internalist state is suitably substituted, then it can have relational antecedents and relational consequences just as externalist states do. We are now faced with two possibilities: 1. The internal facts along with the supplementary material will be sufficient to ensure that the 'subject is in the externalist state in question', or they would not be sufficient. Now if they are not sufficient, then the subject's internal states along with the supplement would 'have all the actual and possible explanatory and counterfactual relations of the externalist state'. And if they are sufficient, then 'the internalist state plus the supplementation is hardly an alternative to the view that externalist states are explanatory'.

3.5 The Explanandum of Psychological Explanation

We can now move on to the other way of countering the internalist charge, viz. by showing that the internalists have made a serious mistake in conceiving the explanandum of psychological explanation. The reason for which the internalists have opted for an internalistic explanation in psychology is (according to externalists) that they have conceived of behaviour (which is the explanandum) narrowly. But this narrow notion of behaviour is not always acceptable. This has

171 'Externalist Explanation' p. 209.
also been demonstrated by Peacocke's above argument. But still we will go into the details of it now.

There is something about which we should be very careful when we talk about narrow behaviour. There is, on the one hand, broad behaviour, and, on the other hand, narrow behaviour. But what is narrow behaviour? Can it be identified with bodily movements? Non-broad behaviour should not necessarily mean bodily movements. One can say that we could have behaviour which is intentional and yet not relational. And so in such cases we are not talking solely about bodily movements, as the same non-relational or quasi-relational narrow intentional behaviour may have two different bodily movements corresponding to it. This is obvious in cases like waving to a friend whom you are currently hallucinating. This behaviour is not wide, as it doesn't involve a relation with an existing object and yet it is intentional and it might correspond to two or three or countless bodily movements, as we can wave at a person in ever so many ways. At this juncture we have to ask the very important question: What is the correct individuation of behaviour, broad or narrow? What could be the grounds for regarding behaviour to be individuated narrowly or broadly?

In order to counter the internalists, the externalists have to show that neither bodily movement nor narrow yet intentional behaviour can be accepted. I shall discuss these two notions of behaviour separately and try to show how an externalist might criticise both these notions.
3.5.1 Behaviour as Bodily Movement

We shall once again be discussing Fodor when we consider the externalist position that the behaviour which we intend to explain in psychology has got to be bodily movement. What we have already said about Fodor is in fact enough to show that his view might lead to the position that behaviour is bodily movement. But still I would show how it definitely does.

From what we know of Fodor already, we can say that it is only because Fodor has conceived of the behavioural consequences of mental states non-intentionally, that he thinks that mental states which give rise to those indistinguishable behavioural consequences have to be the same. We have already argued why it is wrong to think that Ed and Ted’s water-thoughts, water desires, give rise to water behaviour. As we have pointed out in Section 3.3, we should realise that Ed and Ted’s thoughts in one case gave rise to H2O-involving behaviour and in the other case XYZ-involving behaviour. If we thus conceive of the consequent behaviours intentionally (and in this case we consider intentional as broad)172, then we could distinguish the causal powers of Ed’s water-thoughts, water desires from the causal powers of Ted’s water-thoughts, water desires.

But as we have seen in Section 3.3, Fodor is aware of this argument and he has counter-arguments as well. Fodor tries to show why the narrow individuation of behaviour and mental states is more plausible. The point that Fodor makes in favour of this is similar to the one he makes for narrow individuation of mental states. He says that
unless we make a broad individuation of mental states, we will not be able to retain psycho-physical supervenience. It is because of this that one cannot say that Ed and Ted’s *water* behaviours differ. If we do say this, then we have to give up supervenience. And it is absurd to give up the supervenience claim as without this it would be difficult to make sense of how the mental states can at all have causal powers.

Fodor insists that in order to understand his position we have to make a distinction between methodological solipsism with methodological individualism. Fodor thinks that externalists confuse methodological solipsism with methodological individualism. The claim that methodological individualism makes is quite distinct from the claim that methodological solipsism makes. All that methodological individualism maintains is that in psychology mental states have to be individuated in terms of their causal powers. Whereas methodological solipsists claim that we do not need to bring in the semantic properties of a mental state when we need to individuate them in scientific psychology. Under methodological solipsism individuation has to be non-relational. But it is not so in the case of methodological individualism. Individualism has nothing against the individuation of mental states being relational as such. What it wants to emphasise is that whatever property (relational or non-relational) of mental states we consider while individuating them, these properties have got to be such that they can affect the causal power of mental states. Fodor says: ‘Individualism does not prohibit the relational individuation of mental state; it just says that no property of mental states, relational or otherwise, counts taxonomically unless it affects causal powers.’

\[173\] Fodor wants to make a

---

\[172\] We shall discuss later that intentional does not necessarily have to be broad.

\[173\] *Psychosemantics*, p. 42.
methodological point and a metaphysical point. The methodological point is this: The kind of characterisation that we have in science is always made in terms of causal powers. And identity of causal powers depends upon the identity of causal consequences across nomologically possible contexts. The metaphysical point that Fodor is making is that we have to accept mind-brain supervenience, otherwise we would not be able to explain how mental states can have causal powers at all. The causal power of mental states supervenes on their local neural structure. The principle of supervenience also constrains the way that causal consequences are individuated. As a result, the way behavioural consequences are individuated under this principle is different from the way that commonsense psychology individuates it. Under Fodor notion of behaviour which could be regarded as the narrow notion of behaviour, behaviours which are intentional are reduced to mere bodily movements.

Now there are serious problems with conceiving behaviour in this way. We find in Fred Dretske a detailed discussion about why behaviour cannot be regarded as simply bodily movement. In Dretske we find a way of distinguishing movements from moving in the context of behaviour. And he tries to show how the notion of behaviour is closer to the notion of moving than it is to movement. We also find him supporting the view that it is behaviour in the sense of moving rather than movement which is what is being explained in psychology, as it is only moving which can be regarded as behaviour and not movement. To take his own example, there is difference between a rat’s moving its paw and I moving its paw. When I move the rat’s paw, then it is certainly not

174 We will discuss about supervenience and mental causation in detail in the Fourth Chapter.
the rat’s behaviour. It is only when the rat moves its paw that it is a rat behaviour. This is obvious because one could move the rat’s paw even when the rat was dead and in such a case definitely it is not going to be a rat behaviour. The difference between the rat’s moving its paw and I moving its paw lies in the difference of the causes of the two events. Dretske says that if the cause of the movement is in some way inside the animal or person, then it is certainly a behaviour of that animal or person. But it is definitely not so in the case of my moving the paw of the rat. Dretske says that ‘movement that has its causal origin within the system whose parts are moving’\textsuperscript{176} which can be regarded as behaviours of that system. In other words, only those movements of a system which are caused by ‘internal causes’, which can be regarded as behaviours of that system. Dretske points out that by internal cause he does not mean ‘simply inside or underneath the skin’\textsuperscript{177}. This is interesting, as one would like to know what would Dretske consider as internal. And here again we could embark on a discussion of what could be internal to the subject. But we will postpone this discussion to the Fourth Chapter, keeping in mind what we said about this in the Second Chapter while discussing Davidson on Self Knowledge.

Now what an externalist would be curious to know is, whether Dretske takes behaviour close to actions or not. Actions too are done as effects of some internal cause or other: a wish, a resolution, a belief, and a decision. This is interesting for us because beliefs, wishes, may well have broad propositional contents. And so if Dretske were to identify behaviour with actions, then he would have to take an externalist line. However, Dretske thinks that behaviours such as these (i.e. actions) need

\textsuperscript{176} Ibid., p.2
some special kind of internal causes. These behaviours are definitely those which can be characterised as voluntary, behaviours which are results of deliberate intentions. Dretske of course does not want to exclude these voluntary behaviours from the purview of his discussion. But he is more interested in having a kind of general notion of behaviour which could apply to all sorts of behaviours (both voluntary and non-voluntary) in all kinds of systems.

At this point Dretske wants to consider a possible objection to his own view. He says that one might think that reflexive behaviours which are caused by external stimuli cannot be said to be caused by internal causes. But nonetheless these are behaviours. To this Dretske says that he too considers these as behaviour, and here too he thinks there has to be some internal causes which have generated the behaviour in question. Probably the internal causing takes place even before the brain registers that a behaviour is going to be elicited by the system. But none the less it is a behaviour as the system in question does it. Maybe not intentionally, but certainly it is the system which does it. And hence for Dretske, internally produced movements (which are counted as movings) are behaviour, and they do not need to be intentional in order to be counted as behaviours. And it is here that the distinction between behaviours which are not actions and behaviours which are actions comes in. Dretske says that when we look for answers to questions as to why and how certain things are done by animals and other living organisms, it is not always the case that we can find the answer in the intentions. The answer might well lie elsewhere. All that Dretske wants to emphasise is that the answer has to be within the organism in question. According to Dretske, in psychology we are looking for explanations of all kinds of

177 Ibid., p. 3
behaviour. And hence, he feels, it would be better if we find a general way of talking about behaviour, a general conception of behaviour which would take care of all kinds of behaviour in question.

Philosophers have, however, been keener on finding explanations of actions, (i.e. intentional behaviours) rather than behaviours which are not intentional. Can we find a general characterisation of behaviour which would come in handy in cases of intentional as well as non-intentional behaviours of humans and other living organisms? Dretske seems to think we can. I however find Dretske's thesis problematic. The reason why I find it problematic is that I find it difficult to delineate intentions from behaviours. I shall say why.

The biggest problem with Dretske's thesis is that it is difficult to decide which of the movements which are internally caused can be regarded as behaviour without invoking the notion of intention in some way or other. As Dretske himself recognises, not all internally produced movements are behaviours. There maybe many internally produced movements which, by no means, are behaviours. For instance, our hair grows, our teeth decay. These cannot be regarded as behaviours strictly speaking, as we do not do these things. They simply happen to us. So it seems that Dretske wants to exclude movements which an organism cannot prevent from taking place within it from the designation of the term 'behaviour'. Consequently movements over which an organism has some control are what constitute the designation of the term 'behaviour'. So behaviour becomes movements which are internally caused and over which an organism has some control. Can we make sense of such a conception of movement without somehow bringing in the organism's intentions? It seems not. Can we make sense of our moving our hands
without bringing in the notion of our intending to do so, especially if we are to distinguish it from the mere movement of our hands which we could not help. What I find problematic with Dretske’s view is that he wants to have a conception of internally caused non-intentional ‘moving’ (as opposed to movements that merely happen to us). I find it difficult to see how a moving which an organism does (as opposed to a movement that happens to it) can be non-intentional. I do not mean to say that all behaviours which are truly our own can be taken to be overtly intentional. But then again there are behaviours which are distinctively intentional. Dretske himself speaks of them. Dretske refers to Davidson and says that there are certain verbs which describe behaviours which cannot but be intentional (voluntary). For instance: asserting, cheating, lying, stalking, etc. And such descriptions definitely indicate a special kind of internal cause. But Dretske thinks that it would not be quite right to think, as many philosophers do, that almost all behaviours worth the name should have such special causes like intentions, or beliefs, or desires, or plans behind them. And hence we cannot take those descriptions as a rule applying to all sorts of behaviour. But perhaps it would not be wrong to claim that in the case of these behaviours with special causes we do need to individuate them in terms of these special causes. For instance we can describe a person’s behaviour as a ‘disturbing neighbour behaviour’, unless that is what was intended by the subject. And in such a case we must accept that these kinds of behaviours might well have to be individuated externalistically.

However, Dretske would not be quite happy with this proposal. He thinks that we can speak about behaviours not only with respect to humans and other animals, but also plants and machines. And for the science of behaviour we need to have a uniform characterisation of
behaviour which will apply even in the case of plants and machines. He says that even in the case of plants and machines we can speak about their behaviour, as we can distinguish between things that happen to them and things that they do. For instance a vacuum cleaner picks up dirt from the carpet. This is a thing that it does. On the other hand the vacuum cleaner may get rusty with time. This is what happens to it. So in the first case we have a behaviour of the vacuum cleaner, whereas in the second case we do not.

Dretske wants us to notice that it is not always possible to find a unique cause for a particular movement, be it internal or external. Many things might go into bringing about a particular movement. How can we sift out the specific cause from all of these. Behaviours are regarded as movements which have some specific aetiology. A movement would, under this view, be a behaviour only if it is produced by some appropriate internal cause. But here there is a problem. How do we tell merely by seeing a movement if it is a behaviour or not. Movements do not come with their aetiology written on their faces. So what kind of movements would count as behaviour? - That is the question.

Dretske speaks of another problem. The thesis that a movement produced by some appropriate internal cause is behaviour is itself ambiguous. It might mean two different things: '...a movement which is produced by some internal cause' or '...a movement's being produced by some internal cause'. The first identifies behaviour with a movement (like the rat's paw movement) with a particular cause. This is an event. The second identifies behaviour with the bringing about of the movement. Dretske says that we cannot identify behaviour with bodily
movements simpliciter. Then that would amount to conflating the above difference. These two alternatives speak of two extremes. Let us take an example to show how neither of the two extremes can be accepted. Fodor's example is of Booth killing Lincoln. Booth did not succeed in killing Lincoln right away. It took Lincoln some time to die. It is true that until and unless Lincoln died, we could not claim that Booth killed him. But this does not mean that Booth killed Lincoln on the moment of Lincoln's death. Killing of Lincoln began long before he actually died.

This idea has led some philosophers, like Jennifer Hornsby\(^{179}\), to think that it is wrong to see actions in terms of overt movements in which behaviours finally end. Philosophers like Hornsby bring in the notion of trying to make this idea clear. We shall discuss in detail later. Dretske finds this view untenable. He thinks that both these extremes are wrong, because 'they locate behaviour in the wrong place — either wholly after it begins or wholly before it ends'. What Dretske suggest is that behaviour is a process. And he hopes that this view would avoid the paradoxes that the two extreme views led to.

One might find something problematic about this view. If we maintain that the behaviour is a process, we have to say that the behaviour lasts throughout the time till the end movement is brought about. This seems strange because, to take the Lincoln example, we would have to say that Booth was killing Lincoln even when maybe he was just sleeping before Lincoln died. Dretske acknowledges the paradoxical nature of such behaviours. In cases of behaviours which take a real long time to bring about the final outcome, it does seem strange to say that the behaviour lasts as long as it takes to bring about the final

\(^{179}\) Dretske, *Explaining Behaviour*, p. 15.
outcome. And it is for this reason that we seldom use the present progressive tense of the verb in these cases. It becomes even more paradoxical because in certain cases of such behaviours, which might even take years to bring about the end outcome, the agent in question can even die in the middle of the process. At this Dretske points out that such paradoxical results plague even theories of perception. Suppose, there is a very distant star which I have managed to spot through my telescope. Suppose further that the star has gone into non-existence by the time that the light from it has travelled to my eyes and has produced the image of it. In such a case we would be forced to say that a non-existing star has caused me to see it. Dretske suggests that we should think on similar lines when we conceive of behaviour as a process.

But all this is not so important for us. We have to see if we have to identify this behaviour, which cannot be merely bodily movement, be it a process or not, it has to be identified internalistically or externalistically. From what we have learnt from Peacocke, it seems to me, that we have certain cases of behaviours which have to be identified in terms of their relational properties and in those cases externalist individuation is necessary. And it is in those cases that externalist explanation is necessary too.

3.5.2 Narrow Intentional Behaviour

We can now consider another line of internalist thesis which says that we can conceive of narrow behaviour as intentional and yet not relational. First of all let us ask ourselves the question: what is narrow

behaviour? Can it be identified with bodily movements? Internalists think that narrow behaviour should not necessarily mean bodily movements, as we could have behaviour which is intentional and yet not relational. And so in such cases we are not talking solely about bodily movements, as the same non-relational or quasi-relational narrow intentional behaviour may have two different bodily movements corresponding to it. This is obvious in cases like waving to a friend whom you are currently hallucinating. This behaviour is not wide as it doesn't involve a relation with an existing object and yet it is intentional and it might correspond to two or three or countless bodily movements as we can wave at a person in ever so many ways. The internalists by opting for this kind of notion of behaviour would actually be in agreement with the externalists, as they consider the identification of behaviour with bodily movements unsatisfactory for the same reasons as the externalists do. For instance, they might think that what we try to explain in psychology is a richer thing than mere bodily movements. One could have a richer yet narrow notion of behaviour which is non-relational and yet intentional. And being intentional it is more than the mere physical bodily movements. For instance, as we have already said, the same intentional behaviour of waving can correspond to two different bodily movements. And the same bodily movement might correspond to two different intentional behaviours. Though this narrow notion has the benefits of both wide behaviour and bodily movement, yet I find it hard to accept. The advantage of this notion is that as it is intentional it is rich enough to make our folk intuitions intelligible and being narrow it will have all the desired merits of an explanandum of a science which categorises entities in terms of their causal powers.
I however find this non-relational and yet intentional notion of behaviour strange. What kind of pseudo-relationality is involved in this behaviour which makes it intentional and yet not relational? I feel that either the internalists have to make this pseudo-relational character of the intentional behaviour clear or they have to opt for a purely physicalistic notion of behaviour, one which reduces it to bodily, and merely bodily movement. If they do not opt for such a notion, then they have to opt for a relational notion of behaviour which is externalistic. Otherwise they will have to opt for a strange notion of narrow yet intentional behaviour.

The main reason for not finding this notion acceptable is: the general problem that we have with the intentional seems to be plaguing this narrow conception of behaviour as well. Now, what exactly is this general problem of intentionality and how does it affect the conception of narrow behaviour as well? We know that intentionality is the capacity of the mind to direct itself upon things. When we say that mental states have intentionality, what we mean is that they are always directed towards something. The general problem about intentionality is that it forces us to accept three things which cannot be accepted together. They are:

1. Intentionality seems to be like a relation, a relation between the thinker and the thing thought about;
2. A relation between A and B entails the existence of A and B;
3. But intentional states can be about non-existent objects.

It is obvious that all these cannot be accepted together. As mental states could be about non-existing objects, it is difficult to figure out how we

---

can make sense of the 'relation' between the mental state and the
intentional object. The most obvious way out from this problem is to
deny 1. But then how could we make sense of the claim that the 'mental
states are directed towards their objects'? The other option is just not to
give much importance to the third claim. But the whole of the internalist
claim heavily relies on the fact that we can have mental states which are
about non-existent objects, as we have discussed in the first chapter.

This general problem will also plague the concept of narrow
behaviour conceived as non-relational and yet intentional. How can we
make sense of this quasi-relational nature of the intentional?

It seems to me that to have an answer to this puzzle we
have to have an answer to the general problem of intentionality. This is
definitely too large an issue to deal with here. And I am not going into
the discussion of it here. All that is evident from this is that if the
internalists identify behaviour with bodily movement, then they do not
have to opt for such a mysterious quasi-relational notion of intentional
and yet non-relationa l behaviour. But you cannot have the cake and eat it
too! It is like saying that the mental states are not relational or world
involving and yet they are intentional. How can we mark off this non-
relational intentional from the mere brain states which are not
characterised by intentionality? Consequently, the narrow notion of
behaviour and the narrow notion of mental states thus seem to me to be
too mysterious! Possibly, considering all this, it would be better to take
behaviour as identical with bodily movement.

There is another point that can be made in favour of the
externalist conception of broad behaviour. The reason that internalists
found it difficult to accept the externalist notion of broad behaviour is that they assumed that broad behaviour has to involve the actual existence of the objects and kinds upon which it depends. Some of the externalists might have been responsible for this misconception. But this is not what has to be true of broad behaviour as such. It might be that behaviours become broad because they depend upon the existence of the object in the past. For instance water directed behaviour does not require the current existence of water.

But this again might make an internalist say that in such a case the externalist's relational behaviour might well be accused of being equally mysterious, as it involves relations to past objects which might even be non-existent now. May be we can answer this by saying that the present behaviour becomes broad as it is generated by a present mental state which in its turn was causally dependent upon a prior relation that existed between a (then) present object and a (then) present mental state. And this is what distinguishes this broad behaviour from the non-relational yet intentional narrow behaviour.\(^{181}\)

Whatever notion of behaviour we might opt for broad or narrow there seems to be one problem lurking. If we need to place the mind in the physical world of causes and effects and if we believe that the mental does perform the role of a cause in such a world then we have to answer the question: How does the mind cause changes in the physical world? This is the question that we will consider now.

\(^{181}\) I am indebted to Paul Noordhof for this point.
Chapter Four

Externalism and Mental Causation

4.1 Introduction

We have seen from the previous chapter that the balance of arguments is in the favour of externalism. But there seems to be one lurking problem that externalism (as we have found it to be, so far) is yet to solve. This is the problem of mental causation. Critics of externalism have tried to say that if we identify mental states externalistically, or if we take mentalistic content to be externalistically determined, then mental states cannot be causally efficacious. And hence on an externalist view of it, the mental would become causally inert.

In a very general way we can say that any thesis about the mind has to provide an explanation of our popular conviction that mental states can bring about changes in the physical world. For instance, we often say that if we are strong-willed enough, then we can achieve many things. We often mean by this that we can bring about changes in the physical world just by virtue of being thus strong-willed. Now, how can a will to do something cause something physical to happen? We need to explain this popular conviction about the causal powers of the mind; to be more specific, it's ability to causally interact with the physical world. Theories of mind need to provide some kind of explanation of (or even explain away) mental causation. And hence we find most theories of mind addressing the problem of mental causation.
Now what is the problem of mental causation? Though in the most general way we can say that it is the problem regarding how the mind fits into our physical world of causes and effects, yet we could identify two distinct problems regarding mental causation. On the one hand, there is the problem that how is non-physical causation possible. On the other hand, there is the problem about content and causation. When we speak about the problem of mental causation for externalism, which of these two problems are we talking about? Is it the problem of non-physical causation or is it the problem of content and causation? The problem of mental causation in general (i.e. the problem of non-physical causation) is a problem for anyone who wants to say there is mental causation, regardless of whatever they say about content and content causation. As externalism too is a thesis about the mind and the mental, it has to solve that general problem, like any other thesis about the mind and the mental. But in addition to the general problem, there is a specific problem for externalists.

Let us see what this specific problem is. It is a matter of common belief that our thoughts are able to move us into doing things (in the physical sense), because the thoughts have particular contents. Now suppose the contents of our thoughts are externalistically conceived (that is, suppose we were to claim that there are certain mental states whose existence entails that there are objects which they are about or entails that the environment is in a certain way). In such a case we would face the second problem: the problem of content causation\(^\text{182}\). I see the problem of non-physical causation as the broader problem and the

content causation problem as a sub-problem within its fold. Historically speaking, the problem of non-physical causation is prior to the problem of content causation and hence the first provides the second a sort of background. I feel a full understanding of the first is required for the full understanding of the second. And so in this chapter I shall discuss both.

There is another specific and more important reason why I think we could do better if we could settle the non-physical causation problem first. I think this would help us to tackle the content causation problem. What exactly is the relation between these two kinds of problems and their solutions? To make a moderate claim, it seems to me that in understanding one of the problems we can actually come up with a better understanding and finally a better solution of the other. On a less moderate stand, I am inclined to believe that we could have a solution for the non-physical causation problem which would help us to resolve (actually dissolve) the content causation problem. When I say this, I have a particular solution of the non-physical causation in mind. Not any solution to the non-physical causation will give us this bonus, as we have in the literature certain solutions of non-physical causation which could not count as a solution to the content causation problem. The view that I have in mind is Davidson’s proposed solution to the mental causation problem. In Davidson we have a kind of double answer. This he does as he questions a very general way of looking at causation itself. What he suggests is that this general way of looking at causation might have given rise to both the problems. And once we drop this approach to causation in general, we will possibly manage to solve both the problems at one go. This general approach to causation is tagged with the popular belief that causes manage to produce their desired effects by virtue of
having some property. Davidson argues that this belief is wrong, and that if we are ready to give it up, then we can have a solution to the first problem (and have a dissolution of the second problem).

This twin solution is coherent with the spirit of externalism and I will try to show how it is. Of course we know from our earlier discussion on self-knowledge that Davidson is an externalist. And so one can be a Davidsonian and also an externalist. And such an externalist would have in hand a twin solution to the problem of non-physical causation and content causation. In this Chapter, I want to defend this (Davidsonian) option. Davidson's writings are mainly on the problem of non-physical causation and I intend to show how we can get a solution of the content causation problem from what he says in them. To do this we need to discuss in full Davidson's Anomalous Monism. This discussion would show how we could circumvent the whole problem that the internalists posed for externalism.

Before going into a Davidsonian story, let us try to understand the two problems as they arise in the literature and try to understand how an understanding of one could give some insight into the understanding of the other. So we shall first see why it is thought that externalism has a special problem with mental causation and then see what the general problem of mental causation is and how it affects not only externalists but all the others who are seeking to find an answer to how the mind fits into the world of causes. After considering these we shall go on to see how on a Davidsonian view there is an answer to the problem of non-physical causation.
physical causation (provided you accept Davidson's metaphysical views on causation) and such a view is well in tune with externalism. And we shall see, under the Davidsonian view, the problem of content causation does not arise at all. In the end, I shall try to show that even if one were not to accept a Davidsonian view regarding mental causation, still it could be argued that the problem of mental causation is not a problem specific to externalism.

4.2 The Externalist's Problem

What exactly is involved in our common belief that our thoughts can move us? It is not just that we commonly believe that the purely physical/spatio-temporal properties of our thoughts are causally responsible for our behaviour. Rather, it is the content of our thoughts which are responsible in bringing about the behaviour in question. Philosophers have found it difficult to show how contents can bring about changes in the world. This is the problem of content causation and it is a problem for anyone who believes that thoughts manage to bring about changes in the world by virtue of their contents. This is the general problem of content causation. And in addition to this general problem about content causation, externalists have a special problem as they conceive content externalistically.

Let us first try to understand what the general problem of content causation is. It is this: When we speak about contents of beliefs, we are talking about propositional contents of beliefs. And the propositional contents are regarded as abstract and it is hard to see how an abstract

185 One thing is clear that Davidson was trying to come up with a solution to the non-physical causation problem and that actually showed how the content causation
object like propositional content can be a cause. Causal relation is usually viewed as a spatio-temporal relation. And if it indeed is so, then it is difficult to see how the abstract propositional content can be causally efficacious. One can say that though the abstract content of a thought itself cannot cause anything, yet the particular episode of thinking a thought, with a particular content, can certainly be a cause. It is something like saying that abstract things like properties cannot cause anything, but the objects with such properties can surely be causes, by having those properties. It is as though in those properties the causal power of the object is nested. But this analogy between thoughts with abstract contents and objects with abstract properties doesn't hold. This is so because in the case of objects with properties, the objects are regarded as the instance of the properties; but it is doubtful as to whether thoughts are instances of their abstract contents. It is true that the particular instance of the property is not abstract. But this cannot be what is happening in the case of thoughts and their contents. The content remains abstract, it is not as though the thought episode was an instance of the abstract content, because then beliefs and desires with the same content would be instances of the same properties.

There could be a possible solution to this problem. The mere fact that contents of mental states are abstract does not right away make states with content causally inert. We use abstract numbers to measure physical magnitudes like heights or weights, but that does not make weights and heights causally inert even though they are measured in terms of abstract numbers. And weights and measures can definitely be causes. For instance, that we fall to the ground when unsupported, is
partly due to our weight. And such a weight is measured by numbers which are definitely abstract. In this way, perhaps, we could find a plausible answer to the general problem of content causation. But the specific content causation problem that affects externalism still remains.

The specific problem arises because externalists conceive of the contents of thoughts externalistically. We have to remember that how the content is conceived is very important here. This is so because it is commonly believed that an event could not be an event of thinking unless it had a particular content, and in this sense the content was essential to the thought episode. So although the thinking episode can be credited with doing the causing purely in virtue of its natural properties, the explanatory relevance of content of the thought is guaranteed because it is an essential component in the individuation of the thinking episode.\textsuperscript{187} But the real question is: Does the content make the thinking episode causally efficacious? To put it in another way: Is the content intrinsic for the thought to cause changes in the world? Saying that a causally efficacious thing would not be what it is, unless it had a particular content, would not right away make the content itself causally efficacious. It seems to be once removed from the causal efficacy. And, moreover, how can we claim that it is the content which made it causally efficacious? In a way we can make such a claim, provided that we can show, somehow, that the content was intrinsic to the thinking.\textsuperscript{188} This is thought by most philosophers, as they believe that in order for something to be intrinsic for a cause to produce its effect, that something

\textsuperscript{186} See Martha Klein's 'Externalism, Content and Causation', Proceedings of the Aristotelian Society, 1996.

\textsuperscript{187} See Martha Klein's 'Externalism, Content Causation', p. 160.

\textsuperscript{188} Recall the discussion in the First Chapter on Crane and his 'All the Difference in the World' and his view that a thing could be causally efficacious only in terms of its intrinsic properties.
has to be intrinsic to the cause. Consequently it is believed that only if we could show that the content was intrinsic to the thought, that we would also be able to claim the causal efficacy of the content. But this is where all the questions arise. For instance, if the content was external, then it would be difficult to say that it was intrinsic. In other words, if we take beliefs to be relational, then it is difficult to show how they can cause changes in the physical world. But maybe it is not so obvious why the content fails to be intrinsic, if it is taken to be external. We have seen from our discussion of Davidson's view on self-knowledge that even the external can count as intrinsic. If this be so, then we would be able to show how beliefs with external content could cause changes in the physical world. At least the internalist could not say that the reason that we could not make sense of the causal efficacy of external content was that the external content was non-intrinsic to the mental states. They would have to come up with some other reason for discounting the causal role of external content. But one might not accept Davidson's claim that even the external can be intrinsic. In such a case I propose a stronger argument. I intend to show that the common belief that only that which is intrinsic to a cause can be something which is intrinsic for it to cause an effect is not quite correct. Even something extrinsic to a thought can be what is intrinsic for that thought to cause something. In this section I intend to demonstrate this. But first, we have to understand the opposite position in order to overturn it. And so, for the time being, let us assume that it is only something which is intrinsic, that can count as intrinsic for a cause to produce an effect.

The worry about how externalistic content can be causally efficacious seems to be also associated with the claim about the local

---

189 See Davidson 'Knowing one's Own Mind'.
nature of causation (being local of course is being intrinsic). For instance, this is how Collin McGinn sees the problem. He says that causation is 'the same with brains as it is with billiard balls. Their effects depend upon local properties of these entities'. The mental explanation cannot benefit from content, especially external content, as mental explanation like any other explanation has to be given in terms of local causes.

The real problem is that internalists feel that the things which make a state contentful cannot be that which makes it causal. This is so because it is the relation to the object represented in the state which makes the state contentful, and this is believed to be totally different from what makes it causal. And what makes it causal is something which has to be intrinsic to it. And since this relation which makes a state contentful is not intrinsic (in cases where the content is external), it cannot be what makes the state causal.

This is why there seems to be a big problem in making sense of our common belief that we do what we do as a causal consequence of what we think. And it is a bigger problem for those who talk about externalist content. The problem, as far as I see, is not so much to do with the fact that propositional content is abstract. It is the problem of how can an external object, which is not intrinsic to a mental state, be responsible for the causal efficacy of the mental state. (This can be said in spite of recognising the fact that it is the content which makes the mental state what it is).

---

191 McGinn, Mental Content p. 133.
192 See Klein's 'Externalism, Content and Causation', p. 167.
Well, there can be a few general things that can be said in support of the externalists in this respect. One might ask specially to the people who have spoken about local causes that why is it that the 'outside-the-body location of the content-determining feature poses a greater problem than an inside-the-body location?\textsuperscript{193} Martha Klein says that it cannot just be a matter of spatial distance. It cannot be the case that philosophers are actually saying that the causal power in question has to be located exactly where the state is located; because if it was so then the internal properties would pose exactly the same kind of problem as surely they are not located in exactly the same place either. And it seems difficult to decide where should we draw a line between the external and the internal. The same kind of objection can be raised against those who say that since the relation to an external object cannot be what is intrinsic to the mental state, it cannot be what makes the state causal. That is, we can ask why is it the case that only the features which are intrinsic to the state can be causally efficacious.

One thing is clear from the discussion of the internalist position, the words 'intrinsic', 'internal', and 'local' have been used in this connection in a rather sloppy manner: sometimes inter-changeably and sometimes to-mean totally different things. I think that to make sense of what exactly is going on, we have to be clear about what these three terms mean, and how they are related.

To start with, we might try to understand what we mean by a property's being intrinsic. Following some philosophers we might

\textsuperscript{193} See Klein 'Externalism, Content, Causation', p. 167.
invoke the notion of 'Cambridge change'\textsuperscript{194} in order to make sense of what could constitute a property's being intrinsic. Let us take Peter Geach's example: If \(a\) gets taller than \(b\) by outgrowing him, then this is a real change; whereas if \(a\) gets taller than \(b\) as \(b\) shrinks, then it is a mere Cambridge Change. L. L. Humberstone in his paper 'Intrinsic/Extrinsic'\textsuperscript{195} has tried to make this distinction clear. When \(a\) becomes taller than \(b\) by outgrowing him, then 'we can say that the relation (taller than) which comes to hold between them is "reducible" in the sense that there are intrinsic properties ... of \(a\) and \(b\) taken separately, their respective possession of which entails that the relation holds between them'. It is because the change in \(a\), which made him taller than \(b\), is a change of his intrinsic property. This is how the notion of intrinsic property is understood in terms of Cambridge Changes.

It is not quite clear how we can apply this notion to our case. Moreover, there still lingers a question. Is there any non-question-begging-way of saying that we will take those changes as real ones, the ones which are changes in intrinsic properties? We find philosophers coming up with different senses of intrinsic. It is important for us to understand what the internalists mean by intrinsic when they claim that externalist content, which by failing to be intrinsic, also fails to be causally efficacious.

We can find a good discussion on this controversy in Humberstone's 'Intrinsic/Extrinsic'. Philosophers have thought that in order to solve the internalist and externalist controversy on content causation, we have to be clear about what can count as intrinsic. As I

\textsuperscript{194} This terminology was invoked by Geach in his \textit{God and the Soul}, Routledge and Kegan Paul, London, 1969.
have already mentioned in the beginning of this section, I do not share this view with them. What is crucial for solving the problem of content causation is to understand what is intrinsic for a thing's being a cause. I shall try to show how this is a completely different issue. For me what is important is whether externalistic content can count as being intrinsic for a thought's causing a change in the world. I think that being intrinsic to a thought and being intrinsic for a thought to cause something are two different things altogether. And consequently, externalistic content might well fail to be intrinsic to the thought and yet might well manage to be intrinsic for it to cause a change in the world. I shall argue how this is possible.

But before going into that discussion, let us first try to find how the notions of the intrinsic, the internal, the local are related. We have discussed earlier that local and internal seem to fall apart as something could be internal without being strictly local. There is a significant difference in something's being internal and something's being intrinsic. I might well come to have an internal property accidentally which is by no means intrinsic to me. And we have seen from Davidson's discussion on self knowledge that something can be intrinsic without being internal. Consequently these three notions seem to fall apart and it would not be quite correct to use them interchangeably.

Now we come to the more serious difference — a difference which is very crucial for us. The main reason for which internalists reject the externalist thesis, is that the externalistic content cannot be causally efficacious. The reason for saying this is that only the intrinsic can be causally efficacious and external cannot be intrinsic. From this it seems

\[198\] In *Synthese*, 108, 1996.
that only something intrinsic to us can be intrinsic for us to cause something. This is not quite correct. We have to be very careful here and we have to make a distinction between something's being intrinsic to \( x \) and something's being intrinsic for \( x \) to cause \( y \). First let us make a simple point. Not everything that is intrinsic to me is intrinsic for my bringing about some particular event. My being a woman is intrinsic for my bearing a child but not so intrinsic for my making tea. I could well have made tea without being a woman. But this is a trivial point. The more important point is that something might well be intrinsic for my causing an event without being intrinsic to me. To take an example: Suppose I teach in the philosophy department of a college and have been appointed as a member of the examination committee by the Principal. This, i.e., being a member of the examination committee, is not an intrinsic property of mine. But this however is intrinsic for me to cause editing of question papers for the final examination. This property of being a member of the examination committee is neither local nor internal nor intrinsic to me and yet it is intrinsic for me to cause editing of papers to happen. In this case my being related to the college, and being related to the department, my being nominated by the Principal as a member of the examination committee are the factors which determine my being a member of the examination committee. All these are external factors which are determining the property which is intrinsic for me to cause editing of question papers. And hence this property is by no means internal. It is absurd to say that it is local. I come to have this property by virtue of external relations that hold between me and the college and maybe the Principal, etc. So in this case something is being regarded as intrinsic for \( x \)'s being a cause in spite of failing to be intrinsic to \( x \).

\(^{196}\) I am thankful to Amita Chatterjee for this point.
So we see that there are problems with the claim that it is only the intrinsic which can be regarded as responsible for causing something. On the other hand, there are problems in deciding what would count as intrinsic. For some, it is the essential which is the intrinsic. For others, it is the internal alone which can be intrinsic. For still others, even the external can be intrinsic. However, in spite of all these problems, there are at least some clear cases where we all would agree that the extrinsic features cannot be regarded as causes. We can take Dretske's example of a soprano singer to make this point clear. Suppose, the high pitched note of a soprano singer's song shattered a glass. In this case we can, for sure, say that the meanings of the words of the soprano singer were not responsible for the shattering of the glass.\textsuperscript{197} And this is precisely because the meanings of the song are extrinsic to the song.

There could, however, be a slight problem with the soprano singer example. Suppose the soprano singer's song or singing has two effects. The first is that it shatters the glass and the second is that it makes a lady cry. Now surely in the case of the lady's crying, it could well be the words of the song that made her cry. And if the meanings of the lyrics were non-intrinsic in the first case, then they should be non-intrinsic in the second as well. But it seems they aren't, at least to the extent that they manage to bring about a change in the world. So we have a problem here. Either we have to say that it is not only what is strictly intrinsic which can manage to bring about changes or else we have to say that things like meanings of words of a song heard can cause some physical changes in a person. Otherwise, we have to say that the same thing (the meaning of the words in the song) is becoming the
causing factor with the change in the situation. Now let us try to find out why these two cases are varying. It seems to me that the meanings of the lyrics get mentioned in the second case while it is omitted in the first case because the kind of event we are explaining in the first case is totally different from the kind of event we are explaining in the second. In other words the kind of explanation we are asking for in the first case is radically different from the one we are seeking in the second.\textsuperscript{198}

One thing seems to be clear to me in this context, and I would like to emphasise this point here. It is not so important to find an answer to the following kinds of questions: 'Is it what is intrinsic to the cause which can be counted as doing the causing?', or 'Can something external be regarded as doing the causing?', or 'Is only what is local or internal which can do the causing?'. All that we need to find out is what is it that is intrinsic for a particular event's causing another particular event. And the answer to this will vary from situation to situation. To take our earlier example, when the song causes the glass to shatter, it is the pitch that is intrinsic for the song to cause the breaking of the glass. Whereas when the same song causes a lady to cry, it is the lyrics which are intrinsic for the song's causing the lady to cry. And so both an internal and an external property of an event can be intrinsic for it to cause another event. May be the real question is regarding what it takes for a property to be intrinsic for an event's causing another. I have a Davidsonian hunch about this. The answer to this question has to be tied to the kind

\textsuperscript{197} See his Explaining Behaviour, p. 80.

\textsuperscript{198} This is a point similar to a point made by Davidson while explaining how the mental can be regarded as causally efficacious. The question regarding what would count as causally efficacious is always related to causal explanations. And causal explanations are tied to contexts and interests. We shall discuss this in detail while talking about Davidson. And this again demonstrates how Davidson's views regarding non-physical causation can be used in favour of an externalist thesis.
of causal explanation that is sought in the case at hand. And it is only in relation to the interest with which the explanation is sought and given, that we can understand what would count as intrinsic for the cause-event to bring about the effect-event. We shall discuss this again when we come to discuss Davidson’s anomalous monism.

From all this it follows that the internalists cannot say that the externalists have a specific problem with content causation because they conceive the content externalistically. They can no longer say that being extrinsic the externalistic content fails to make the thought causally efficacious. They can no longer say that an externalistic content can never be that which is intrinsic for a thought’s bringing about a particular effect.

4.3 Mental Causation

Let us start from the very beginning. Let us try to trace the historical origins of the problem of non-physical causation. The main questions that made us wonder about how non-physical causation was possible are: How can we place the mind in the physical world of causes and effects? What is the relation between the mind and the physical world? Is there any exclusive place for the mind in our world? Is there any significant relation between the mind and the material world? — All these questions became more pressing in the background of René Descartes’ dualism. It is because Descartes conceived of the mind as totally non-physical/non-material, that philosophers found it extremely difficult to fit the mind in the physical/material world. It was in order to emphasise the uniqueness, independence, distinctness of the mind (and its difference from whatever is physical/material) that dualism was
proposed. But do we really need to conceive the mind as thoroughly non-physical/non-material in order to show its uniqueness or distinctness? Could it not be the case that the mind is actually a part of the physical/material world, but it is different from all other things in the physical/material world? This thought could have led to physicalist/materialist theories of the mind. But which ever way we go, the dualist way or the physicalist/materialist way, we will have to find a solution to the problem of mental causation, as both dualism and physicalism/materialism faces this problem, though each in different ways. Let us see how both of them faces this problem.

The problem with Dualism is that even if it manages to establish the mind as a unique object, yet it will have problems in showing how such an object brings about changes in the physical world. But this is something it does need to explain. Who can ignore the popular conviction that the mind and the thoughts, beliefs, desires, fears, hopes nested in it, can bring about changes in the world? Suppose while writing this I feel cold. This leads me to desire a hot cup of coffee. I get up and start a complex set of physical movements required in making coffee. So we see that my feeling of cold and the desire it produced led to a series of changes in my physical states. Now if the mind and the mental states (experiences and attitudes) manage to bring changes in the physical world, as we have seen above, then can they be totally different from whatever is physical? If the mental can bring about changes in the physical world, then there must be some kind of relation between the mental and the physical. And in order to be so related the mental cannot be totally and absolutely different from whatever is physical. It is very difficult for a radical dualist to explain how the mental can come into some kind of relation with the physical. We know from the history of
philosophy that Descartes faced tremendous problem just to explain what the relation between the mental and the physical was.

May be we can circumvent the dualist problem by giving up dualism and opting for some sort of identity theory: If we take the mental to be identical with the neurophysiological, then there would be no problem in showing how they could possibly be related. We also get a simpler explanation of our commonplace happenings like my desire leading to my coffee-making physical movements under this neurophysiological explanation. Neuroscience would say that the cause of my coffee-making physical movements is the brain (neural, physical, chemical) states that immediately precede it. This explanation is a simple one and is less mysterious. This is the attraction of physicalism/materialism. We find a widespread desire amongst commonfolk as well as philosophers to bring everything under the purview of a neat, coherent and well-organised science. And hence if we can bring the mind under such a neat science, then we are happy. This is why the neuroscientific explanation of our commonplace happenings are often more acceptable. So far so good!

We however soon realise that the explanation that neuroscience gives us is flawed in one respect. What the neuroscientific explanation really boils down to is an explanation of physical causation and not mental causation. As under this explanation we are showing how one physical event can cause another physical event. But we do feel the need of mentioning mental events while giving a causal account of why a physical event took place. If you ask me why did I walk towards the ice cream van, I would say that I was moved by a strong desire for ice cream. Though we are definitely attracted by the neat story that
neuroscience provides us, yet it seems that most of our commonplace understanding of commonplace events and their causes gets lost under such a neat explanation. And so we see that the problem of mental causation equally affects the dualist and the neuroscientific solutions.

One may say that the problem for neuroscience and dualism would vanish if we subscribed to a third view which said that the same physical event could be caused by two kinds of causes: the physical cause (viz. the neurophysical state of the brain that precedes the physical event) and the mental cause (viz. the desire or belief or any of that sort of propositional attitudes that rationalises it). This really seems to keep us all happy. But this surely is a false sense of happiness. We would have to countenance systematic overdetermination of behaviour by the mental and the physical. This is implausible.

If we reject overdetermination in the world of causes and effects and if we believe in the Completeness of Physics assumption, then we will have to say that the mind and the mental are causally inert. This thesis is known as Epiphenomenalism. Epiphenomenalism is the thesis that denies mental causation. It is the view that our mental lives exercise no causal influence on the course of physical events. What the epiphenomenalists feel is that it is impossible that mental phenomena can affect what happens physically. This is so because ‘every physical outcome is causally assured already by pre-existing physical circumstances; its mental antecedents are therefore left with nothing to contribute’\(199\). This kind of Epiphenomenalism\(200\) follows from the

\(200\) Yablo, 'Mental Causation', p. 245
exclusion principle, and the theses of physical determinism and dualism. The argument runs like this:

1. The exclusion principle says that if something, say X, is proved to be causally sufficient for something else, say Y, then something, say X*, can be causally relevant to Y only if it is identical with X.

2. Physical determinism says that every physical event Y is such that there is some physical event X which is causally sufficient for it.

3. Dualism says that given a physical event X and a mental event X*, X is distinct from X*.

And from the above three assumptions the epiphenomenalists conclude that mental phenomena are, with respect to physical events, absolutely causally inert.

Epiphenomenalism can pose a threat to all those who believe that there is mental causation in two ways. It can either say that the mental is totally causally inert, or it can say that if the mental can be causally efficacious then it is so only in virtue of some physical characteristics and not in virtue of some mental characteristics. Keeping this in mind and keeping in mind the distinction between types of events and tokens of events, we can have two kinds of epiphenomenalism:

'TOKEN EPIPHENOMENALISM' Mental events cannot cause anything.

'TYPE EPIPHENOMENALISM' No event can cause anything in virtue of falling under a mental type.'

---


Let us begin with the difference between Type Epiphenomenalism and Token Epiphenomenalism. When we refuse to give mental states or events any causal status then we can say two things. We can say that mental events are never causes. Alternatively we can say that mental events can act as causes only in terms of some physiological properties. It has been usually thought that a cause manages to be a cause only by virtue of being endowed with some property. This is how type epiphenomenalism has to be understood. Token epiphenomenalism is said to entail type epiphenomenalism. If we think that an event can cause something by virtue of being mental, i.e. by virtue of falling under the mental type, then the event would be a cause and would be mental. And so if this was so then token epiphenomenalism would be wrong. Consequently if we say that mental events cannot be causes at all then they would not be causes in virtue of a mental property either. This is how token epiphenomenalism entails type epiphenomenalism. However we have to realise that the denial of token epiphenomenalism does not entail denial of type epiphenomenalism. Under type epiphenomenalism we can think of a mental event which is a cause, but only in virtue of having some physical property then such an epiphenomenalism would be opposed to token epiphenomenalism.

Against the backdrop of epiphenomenalism, I shall try to make a survey of the mental causation debate as we find it today. There must be a way in which mental causation can be understood. There is a very easy answer to this. The problem of fitting the mental into the world of cause and effects, which seems to be the physical world, can be solved if we can manage to show that the mental is physical. And so one of most

---

203 We shall discuss the legitimacy of this claim in this chapter.
obvious responses to epiphenomenalism is the thesis that mental events are effective not by way of being mental, but by way of being physical. This of course would lead us to adhering to Type Epiphenomenalism but not to Token Epiphenomenalism. What we are saying is that the mental events which we took to be causally efficacious are indeed causally efficacious. But their causal efficacy is not due to their mental properties but due to their physical properties. No events can count as causes by falling under a mental type or by virtue of having a mental property.

Whenever we allow ourselves to speak in this way, we are adhering to a specific way of seeing causality: Causes have their effects in virtue of some of their properties. This is a view which seems to be intuitively appealing. We tend to think that something manages to cause something else having been endowed with some special causing property. For instance I could say that, that person makes me angry easily because he has this irritating character.

Now the question arises: What are the properties which make mental states able to cause? Given the epiphenomenalist charge it is difficult to see how mental properties manage to be causally efficacious. The only way we can say mental states become causally efficacious is in terms of their physical properties. Physicalists have opted for this solution. Having subscribed to property/type epiphenomenalism, they have managed to restore the causal efficacy of the mental staying within their physicalist framework and also keeping in tune with their Completeness of Physics assumption. Now is this option acceptable? The

---

See Tim Crane's 'The Mental Causation Debate', *Proceedings of the Aristotelian Society, Supplementary Volume*, 1995..
problem with this solution is that the causation that this solution speaks of does not seem to be mental causation worth the name.

In fact physicalism was evolved partly in order to solve the problem of mental causation, i.e. the problem of how mental causation is possible in the physical world. But what is paradoxical is that physicalism, as it stands now, finds it difficult to explain mental causation, although originally it was conceived to handle the problem of mental causation. There are two problems associated with this physicalist approach. To start with, if the only way that we could make sense of how the mental can be causally efficacious is by making it physical, then we are bound to have a problem, as the same mental state can be realised by different physical states. For example, the same Turin Machine can have various physical realisations. This is known as the multiple or variable realisation thesis. We have to however remember that this does not mean that any and every physical realisation can be regarded as a genuine realisation. There are restrictions which will indicate which of the various realisations are genuine. But even then we can say that there are multiple ways in which such realisations can be made. Well one could take this as an empirical thesis which claims that there are certain mental states which are in fact realised by many different physical states. Alternatively one could take it as a thesis which speaks of a theoretical possibility of the same mental state having variable physical ways of realisation. This definitely makes the mental causation problem a problem for the physicalist too. I shall not be speaking about the variable realizability problem in detail here. My concern is with the other problem that I claim that some physicalist

---

205 Crane's 'The Mental Causation Debate'.
theses (which have tried to say that one could solve the problem of mental causation by saying that the mental events become causally efficacious by virtue of being physical or by virtue of having physical properties) face.

My motivation here is to find out what is it in physicalism that makes it especially susceptible to the problem of mental causation. I will only briefly speak of the various kinds of physicalist theses as I want to sift out the basic claim of most physicalist theses which could be responsible for their failure to solve the problem of mental causation. I will try to find out what the presuppositions that most physicalist these have which might have been at the root of all these problems. We can sift out these presuppositions from the arguments that the physicalists have given in favour of their thesis. And so in the next section I shall discuss the different kinds of arguments that physicalists have given in order to detect the malady nested in them.

4.4 The Problem for Physicalism

Physicalism is roughly the thesis that the mental is the physical. And by physical it is usually meant whatever can fall in the purview of physical sciences. So physicalism is actually speaking about some sort of identity between the mental and the physical. So let us see what the different kinds of Identity Theories have to say about the mental:

1. Eliminative Materialism
2. Behaviourism
3. Type and Token Identity theories

A look into the history of functionalism will make this clear. See 'Functionalism 1-2' in Companion to the Philosophy of Mind, pp. 317–328.
What I want to demonstrate here is this: These theories have a strange problem. Though these have tried to solve the problem of mental causation by opting for some kind of identity between the mental and the physical/material, yet it is their efforts to do this which have made them even more susceptible to the problem.

We should however keep Eliminativism totally apart, as on this view there is no such thing as a mind or the mental even conceptually. Such a thesis doesn’t have to face the problem of mental causation as it does not believe in it. It is only those theories who do believe that the mind and the mental (as, at least, conceptually distinct from the physical) can bring about changes in the physical world that face the problem of mental causation. And so Eliminativism is of no interest to us.

Behaviourism came into focus in the first half of the twentieth century. One could say that behaviourism came about as philosophers and psychologists became wary about the plausibility of introspection as a scientific method in psychology. They were also very impressed by the fact that we seem to be able to have knowledge about other minds by virtue of knowing the behaviour of those others. This gave a sort of epistemic justification for thinking that we could know about mind in general (both our own minds and those of others) from the behavioural clues that are easily accessible. For the behaviourists it is far more
plausible to take the mind as that which is essentially constituted by its outer effects, than to conceive of it as some inner psychic mechanism. Behaviourism was opted for as it supposedly gave us a better methodology than its preceding theories which took introspection as the means of knowing the mind and whatever is psychological. The introspective method (being distincively subjective) was found to be unsuitable a method for scientific theory. Behaviourism provided us with a way of identifying mental states in purely physicalistic terms and hence was believed to be more conducive to a neat science of the mind and the mental. A science of the mind needs to be objective and must help us to make prediction on the basis of observation, and so behaviourists contended that observation of behaviour alone could provide the necessary data for scientific psychology. Consequently some behaviourists wanted to redefine all the relevant terms in psychology behavioristically. But what is behaviour? Ordinarily it is any movement of an organism. But this definition is not satisfactory, as it would then bring in things like the heart's pulsation or the movements in the organism associated with digestion into the purview of psychology. And so philosophers restricted the notion of behaviour to whatever an organism does in order to act upon its environment, i.e. the outside world. And this certainly is observable. But here again we have two sets of problems. Firstly what is observable is not always a good clue to the real characterisation of the behaviour. Suppose you see me writhing in pain and conclude that I am in the psychological state of pain; all the

\[207\] All the other kinds of physicalist thesis would try to avoid conceptual or methodological reductionism. We shall speak about this when we speak about reductionism.

\[208\] This reminds us of the kinds of reservation that philosophers and psychologists nowadays have against invoking 'folk' notions like beliefs, desires and such propositional attitudes in psychology as they think that they are unsuitable for a science of the mind)
behavioural clues indicate that I am indeed in pain. But it could well be that I am faking it altogether. Suppose while pretending to break the glass top of my table with a heavy paperweight, by accident the paperweight slips from my palm and breaks the glass top of my table. Seeing what happens outwardly you would perhaps conclude that I was breaking the glass with the paperweight. It seems that it would be difficult to make a distinction between slipping and deliberately throwing in such cases. In the case in hand I simply wasn't breaking the glass, as I did have no such intention at all. Though it is possible for you to observe my outward behaviour, yet it is impossible for you to observe my intention. The second problem with this picture is that we often have mental states which do not manifest themselves in any kind of behaviour. For example, being in a bad mood is certainly a mental state. But if there is a man who is in full control of his temper, then it would be difficult for a behaviourist to detect that the man was indeed in a bad mood as he would show no behavioural signs of it.

This had led behaviourists to speak about dispositions, which however led them into even deeper troubles. Let us take an example to see how behaviourism sees mental states. Suppose I believe that my bookshelf needs to be organised. This certainly is a mental state and usually this belief would lead to my bookshelf-organising behaviour. The behaviourist would want us to identify my believe that I ought to organise my bookshelf with my disposition to organise the bookshelf. The difficulty with dispositions is that they seem to be once removed in the chain of causal sequence. And dispositions would fair no better in a science of the mind than the introspectible subjective states.
However whether or not behaviourism can survive is of no concern to us. What we find from the above discussion is that if we take behaviourism to have some promise for us in solving the mental causation problem, then we are definitely mistaken. Only an eliminative behaviourist could have an answer to the problem, in which whatever is and was regarded as mental gets replaced by propensities to act physically. But just as plain eliminative materialism could not be our concern, so also eliminative behaviourism; as for such a thesis the problem of mental causation is not a problem at all.

Let us now talk about identity theories. As I have already said, these theories seem to offer the easiest solution to the problem of mental causation. By identifying mental states with physical states, the whole puzzle regarding how we could fit the mind (which was so very different from whatever physical) in the physical world vanishes. If mental states were in fact physical states, then it wouldn’t be difficult to show how they could cause other physical states. But things are not as simple and rosy as it seems. To start with, what exactly it is to say that the mental is identical with the physical. Descartes’ dualism hasn’t lost its appeal so much that identity theories would seem obvious to us. However much it might seems scientifically plausible to us that the putative mental states (such as desires, thoughts, and beliefs) were actually neuro-physiological states, it is difficult to accept that conceptually these mental states were the same as their neuro-physiological counterparts. But then the identity theorists would say that the identity thesis did not rest upon the conceptual reduction of the mental to the neuro-physiological. One could well avoid that and yet speak of identity between the two. Following Leibniz, they could say that all that is required to demonstrate that the mental was identical with
the physical was showing that they shared all their properties. And that would do the job even if they were conceptually different. So we are actually saying that the mental shared all its properties with the physical. But this is not a very easy thing to demonstrate either. The properties which we usually take to be properties of the mind and mental seem to be quite distinct from all our known physical properties. We have in mind both propositional attitudes and sensations. Especially in the case of sensation it is difficult to say how those properties can be understood physicalistically. So what are the properties of the mental like? Are they physical or 'mental' or neutral? Well, we are not going into a detailed discussion of this problem here, as it is not our prime concern. All that is clear from this is that it isn't obvious why the mental should be identical with the physical.

We know of two distinct ways in which we can take the thesis of identity between the physical and the mental. We might speak of identity at the level of types or at the level of tokens, and hence the type identity theories and token identity theories. If a mental state is type identical with a physical state, then all the instances of that particular type of mental state would be actually instances of some particular type of physical state. Token identity theories, on the other hand, would just say that each mental state is identical with some physical state of some type or other.

There are serious problems with type identity theories. We could think of two species which share some of the types of mental states without sharing any types of physical states. To take an example, pain may be a common type of mental states among many kinds of life forms.

---

209 Refer discussion on Qualia and phenomenal states in Chapter Two.
in spite of the fact that those life forms have no physical state types in common. This is known as the multiple realisability thesis about the mental.

There is another problem with the type identity theories. Under the type identity theory a mental state type is identical with a physical state type. One way of understanding this is that the identity theories identify one property with another. This way of looking at the identity between the physical and the mental is inspired by the kind of identity theories we get in physical sciences. To take the popular example from the literature, lightning is identified with electrical discharges generated by clouds, because lightning is regarded as the one and the same property of electrical discharges generated by clouds. But we have to keep in mind that the identity between such properties is guaranteed by bridge laws linking the two properties. And this is what is missing in the case of mental and the physical. We can have no bridge laws linking the mental and the physical. If we could indeed find such bridge laws linking the physical and the mental, then we could reduce the mental to the physical. In the absence of such bridge laws none of the reductionist theses would be acceptable.

There are other problems with reductionism. Critics say that it is impossible to find a physical property, no matter how complex it is, which could be identical with mental property. But one could just give up identity at the level of properties and stick to identity at the level of entities. This gave birth to aspect dualism or property dualism which can go hand in hand with token identity theories. Davidson is an advocate of this kind of identity theory. We shall discuss this brand of non-reductive
token identity theory later in detail. But for the time being we must see why type identity theories cannot be accepted in the absence of bridge laws and in the presence of the danger of multiple realisability of mental states.

Functionalism too, like all the theories we have been discussing, tries to find out what is it that makes a mental states or event mental. Under Functionalism, a mental state or event is the particular kind of state/event only because of the functional relations that it has with the subject's perceptual stimuli, behavioural responses and other mental states. Under this view mental states can be defined in terms of their causal connections. We have to see the mind, according to functionalism, in the way that we see carburettors or kidneys. All these are functional concepts. If we ask 'What is it for a thing to be a carburettor?', the answer we get is 'To be a carburettor a thing has to mix fuel and air in an internal combustion engine'. We can answer similar questions about the mind in the similar way, on functionalist thesis. If we were to ask a functionalist, 'What is it be in pain?' she would answer, 'Being in pain is nothing but having some property that makes you groan or bring tears in your eyes, or makes you think you must be ill'. So we can see mental states just like the functional states of a machine. But the same functional state can be realised in very many ways. We can see being a typewriter as being a thing that writes, but such a function can be realised by ever so many ways. And functionally all of these would be the same. And this would be true of mental states too if they were seen functionally. This has made functionalism competing with physicalism. We can say that distinct species with noticeable anatomical and physiological differences can exhibit same functional states. And hence they would be functionally

210 We shall be discussing about bridge laws again, later in the chapter.
similar and yet be physically different. And hence there is a possibility of
the same functional mental states being realised in distinct physiological
systems. This would lead to an asymmetry between functionalism and
physicalism. Such a functionalist thesis would say that the same mental
state can be realised in my different physical systems and it is even
logically possible that they are realised by non-physical systems. And so
sameness and difference of such mental states would not be tied with the
sameness and difference of physical states. Some philosophers would
however say that it is wrong to think that physicalism and functionalism
are opposing theses. Physicalism can address two kinds of questions:
tonological and metaphysical. 211 Ontological physicalism is concerned
with the question if at all there are mental states. Metaphysical
physicalism is concerned with the question as to ‘what gives each type of
mental state its own identity’212. It is this metaphysical physicalism
which is actually type physicalism which is in conflict with
functionalism.

One interesting thing about functionalism is that it is not
obviously opposed to externalism. Under a functional theory one could
talk about ‘long-arm’ inputs and outputs which would include external
objects. Such a functionalist thesis would speak of wide as opposed to
narrow content and could be externalistic and would hence face all the
problems regarding mental causation that externalism faces. But of
course there are internalistic functional theories which take inputs and
outputs as ‘short arm’ which would not include external objects.

211 See Ned Block, ‘Functionalism 2’ in A Companion to the Philosophy of Mind, pp. 325-
326.
212 Ibid., p. 325.
Functionalism too faces the problem of mental causation in its own way. On a functionalist thesis, mental properties are regarded as second order properties. Having a mental property is having other properties which have relations to one another. There are problems in seeing how second order properties can be causal in the way that mental properties have to be causal. We can take Ned Block's example to show what the problem is with secondary properties. We know that a red cape can provoke a bull. So provocativeness in a second order property which consists in having a first order property like redness, which provokes a bull. But what caused the provocation in the bull? Is it the provocativeness of the cape or the redness of the cape? Certainly it is the redness of the cape. So how can the second order property of provocativeness cause the bull to be provoked? The same sort of problem will arise in the case where we see mental properties as second order properties.

From the above discussion we have found out how each of the physicalist schools of thought specifically face the problem of mental causation. Now I want to point out certain features in physicalism (taken as a general thesis) which has led them to face the problem of mental causation. Following Tim Crane's discussion in 'The Mental Causation Debate', it can be pointed out that there are certain presuppositions that almost all kinds of physicalism are adhering to and it is the adherence to all these presuppositions together which have led them to the problem of mental causation.

Let us see which are the commonly held presuppositions which have led us to have the mental causation problem, especially for the physicalists. To start with, we have to remember that we are only concerned with physicalists for whom the mental causation problem is a genuine problem, and consequently we will keep eliminativism out of our discussion. So we can say that the first presupposition of all the physicalist theses that we are concerned with is that there is mental causation.

The second presupposition is the one that we have spoken of in the beginning. Other than Davidson’s anomalous monism, all the other physicalist theses adhere to this presupposition. It is the commonly held belief that causes have their effects in virtue of some of their properties. Now why do we say that all kinds of physicalist thesis adhere to this presupposition? All kinds of physicalist theses, that we are concerned with, try to make sense of the fact that mental events can manage to cause physical events by saying that they do so because they are made causally efficacious by some physical properties that they are endowed with. And being so endowed with physical properties, those putative mental states are in a way physical. This seemed to the physicalists the easiest solution of the mental causation problem at hand. But I think that it is this easy solution which has taken physicalism deeper into the problem of mental causation. If in order to prove the causality of a mental state we have to understand its causality in terms of some physical property (which is actually responsible for the causal efficacy of the mental state), then we cannot term that causality as mental. And so the solution is not a solution of the mental causation problem. So the

214 In discussing the assumptions I am following Tim Crane’s discussion in ‘The Mental Causation Debate’.
causality of the putative mental is no longer mental, it is physical. And so this is unacceptable. This is what can be said generally against this presupposition. But, of course, we remember from our discussion on type identity theory that this would also lead to the problem of multiple realisability. So this presupposition is already a suspect. Later in our discussion we shall see the specific problems that Davidson finds with this presupposition. However, in all fairness of things we should remember that this is a common way of looking into causation and not just one associated with physicalism. It is only in Davidson that we find a radical shift from this common understanding of causality. And this we shall discuss in detail later.

We have already spoken about the biggest assumption on the part of physicalism: the completeness of physics. Now what is the completeness of physics assumption? This assumption says physical science is explanatorily adequate. That is, given any physical phenomenon there is a body of scientific theories which are adequate to give a true and exhaustive physical explanation of it, in principle. And since the physical effects of the putative mental causes are none-the-less physical, they can be exhaustively explained by physical science in terms of purely physical phenomena.

Given this and the belief that causes have their effects in virtue of their properties, the conclusion that follows is that the mental properties are physical properties. This is so because if the putative mental state gives rise to a physical effect, then it must do so in terms of some properties. If we have to prove that mental causation is possible, then we have to say that it is the mental properties which make the mental states
causally efficacious. But then by the completeness of physics hypothesis we should admit that there is a complete physical explanation of the physical effect of the putative mental state. This explanation should be in terms of the physical properties of the mental state in question. If the physical effects did not have physical causes, then they would not be explicable in physical terms and then the completeness of physics assumption would have to be given up. So if mental properties have to be characterised by their causal roles, then they must be physical. At least the mental properties which have physical effects have to be physical.

One might say that there could be two alternative causes of the same effect. But this is one thesis that has been rejected by everyone. And this is one of the other assumptions that have played an important role in the mental causation problem. The view against the possibility of having two alternative causes of the same effect is known as the over-determination argument. The over-determination argument says that every physical effect has a complete physical cause in terms of which it can be explained (as they do believe in the explanatory adequacy of physical science). Now if we accept that in addition to all the physical causes an event might have a mental cause, then we have to say that the physical effect is over-determined by both the mental and the physical causes. But such an over-determination is highly unlikely. And so we have to accept that the so called mental cause is actually identical with the physical cause.

215 See Papineau 'Why Supervenience?'
216 An argument that has been defended by philosophers like Hopkins, in his 'Mental States, Natural Kinds and Psychophysical Laws', Proceedings of the Aristotelian Society,
But why is it that over-determination is regarded as false? In over-determination it is not that the two causes are being seen as partial causes for the effect. The mental cause and physical cause are taken to be independently sufficient for the physical effect. And this seems to be unacceptable. Those who believe in the counterfactual analysis of causation find it very hard to accept this. If X is the cause of Y, in a counterfactual situation if X doesn’t occur so will not Y. And so if both the physical and the mental are regarded as independently causes of the physical effect, then in a counterfactual situation the absence of one ought to entail the absence of the effect, but it does not. So we see that adherence to the completeness of physics assumption along with the denial of overdetermination would make physicalist theses again say that the putative mental becomes causally efficacious by being physical in some way or by having some physical property. This again would make the causation in question physical causation and not mental causation.

There are other problems with the completeness of physics presupposition. We find a detailed discussion on this in Patrick Suppes’ ‘Davidson’s Views on Psychology as a Science’\(^\text{218}\). In this article Patrick Suppes is giving us instances of theories in"physics which study systems that are not closed and hence we cannot claim about these theories in physics that they are complete in the sense we needed them to be. In other words, they might quite well be explanatorily but not adequate, as


they are not deterministic. And so under such a theory it might as well happen that we cannot pick out a physical cause of a physical effect. And yet this would count as a physical effect. And in such cases one could find a non-physical cause of those physical effects.

In view of all this one could say that maybe it is wrong to suppose that mental causation should work in the same way as physical causation does. But surely, causality has to be the same everywhere, it is a general thesis about the way of the world and it cannot be different in different cases. Moreover if the causes are competing, then they must be of the same kind.

Now where exactly does Davidson fit in this picture? Davidson also adheres to most of the assumptions that we have spoken of. But he doesn't adhere to the general way of looking at causality that is nested in the presupposition that causes manage to produce their effects by virtue of having some properties. And this is his way out from the problem of mental causation. We will now go over to the discussion of Davidson. Davidson too intended to show that there was an identity between mental and physical states. But though like all other forms of physicalism this thesis too endorsed 'ontological reduction', it eschewed 'conceptual reduction'. We find Davidson's major thesis in his essay 'Mental Events'220 where he expounds his famous thesis of anomalous monism.

---

219 See Davidson, 'Thinking Causes', in Heil and Mele (ed.), Mental Causation.
4.5 Davidson’s Anomalous Monism

To start our discussion, we have to remember that Davidson does believe that there is mental causation. This he says clearly in ‘Mental Events’. He says that there are at least some mental events which do indeed cause physical events. The crucial question is how is his belief in mental causation consistent with his anomalous monism? One might think that having opted for a token identity theory regarding the mental and the physical, Davidson would be forced to be an eliminative physicalist. And this would make it difficult for him to show how mental states are causally efficacious. The critics think, his problems multiply as he holds that there can be no strict laws connecting the mental and the physical and as he believes in the nomological character of causality. Let us see how all these problems can be circumvented, staying inside the Davidsonian framework.

At the heart of Davidson’s physicalism is his belief that there can never be strict psycho-physical laws. In other words there can be no strict laws on the basis of which mental events can be predicted or explained. One thing should be said right at the beginning: In understanding Davidson’s thesis we have to keep in mind his particular way of looking at the causal relation. He takes the causal relation to be a law-like relation between two particular events. In understanding his thesis regarding causation we also have to keep in mind how he sees events. His conception of events is more a result of a semantic theory.

---

221 See p. 208.
222 We shall discuss how with the help of supervenience relation Davidson can avoid eliminativism.
rather than a result of a strictly metaphysical theory. Davidson says that two events are the same if they have the same causes and effects. Davidson wants us to note that events can be described in very many ways and such varied descriptions need not be logically equivalent or synonymous. We have to keep this in mind when we try to understand what he says about psychophysical laws.

Davidson believes in the nomological character of causality. By this he means that if two events are causally related, then there must be a strict law under which they can be subsumed. This is known as the 'nomical subsumption' thesis. Now what does Davidson mean by a strict law? Before going into strict laws we must understand what Davidson means by laws as such. For him laws are linguistic, as he takes laws to be true lawlike sentences. And so they are parts of the language. And a sentence is lawlike if and only if it is a generalization that is confirmable by its positive instances and which, if true, would support counterfactuals and other subjunctive conditionals.

What does Davidson mean by nomic subsumption? An event is subsumed by a strict law only if it instantiates that law. And Davidson says that events can instantiate laws only relative to certain descriptions. We have to remember this while we speak of nomic subsumption. As

---

223 As Kim says in 'Events as property exemplifications' in his Supervenience and Mind, (Cambridge University Press, Cambridge, 1993) the metaphysical theory of events should be distinguished from the theory of the "logical form" of event and action sentences. What we get from Davidson is a consequence of his semantic thesis regarding action or event sentences.

224 See Davidson's 'The Individuation of events' in his Essays on Actions and Events.

225 See page 208 of 'Mental Events'.

226 See Brian McLaughlin, 'Anomalous monism and the Irreducibility of the Mental' in Le Pore, & McLaughlin (ed.) Actions and Events, p. 333.
Davidson says 'The principle of nomological character of causality must be read carefully: it says that when events are related as cause and effect, they have descriptions that instantiate a law. It does not say that every true singular statement of causality instantiates a law'\textsuperscript{227}. So laws can subsume events only under certain descriptions. When can we say a law covers a case? A law covers a case under this condition: Suppose the law is the sentence $\alpha$; further suppose that $\beta$ is a sentence that says that the cause event (appropriately described) has occurred; then $\alpha$ and $\beta$ would entail a sentence $\chi$ asserting that the effect event (described appropriately) has occurred. We cannot formulate such laws with the mental states and their physical effects. Thus there are no psychophysical laws. Davidson says that there can be heteronomic laws linking the mental and the physical and not homonomic laws. Homonomic laws are strict laws whereas heteronomic laws aren't. Davidson says that if we have a generalisation (and we know that laws are generalisations) which are such that when they are instantiated they may give us reason to believe there is a precise law at work (but one that can be stated only by shifting to a different vocabulary), then that generalisation is a heteronomic one. The kinds of generalisations we have with mental and physical events are exactly like these. We realise that there must be a 'precise law' linking them, but in order to spell it out, we have to shift to a physicalistic vocabulary. We need to do no such thing in cases of relations between physical states and hence the laws that connect them are strict and hence homonomic.

If all this is true, then how can we make sense of the causal connection between the mental and the physical? Davidson reconciles this apparent contradiction by saying mental events are physical events,
ontologically. For if mental and physical events interact causally, then they have descriptions under which they instantiate strict laws. And in order to fall under a physical description (and hence instantiate a law), a mental event has got to be a physical event.

Critics have thought that having spoken of the anomalism of the mental (i.e. having said that there cannot be any ‘strict deterministic laws on the basis of which mental events can be predicted and explained’), and having spoken of the nomological character of causality, Davidson has failed to show how some mental states can be said to interact causally with some physical states. But Davidson has tried to show how all this can fit together.

The anomalism of the mental does not make mental events causally inert. It only says that there can possibly be no strict laws on the basis of which we can explain or predict events under mental descriptions. But events which have mental descriptions and hence can be called mental events can also have physical descriptions under which they manage to instantiate laws. Obviously this presupposes that such ‘mental’ events are identical with physical events. Davidson wants us to understand that ‘we cannot conclude from the fact that \(a\) instantiates a strict law, and the fact that \(a = b\), that \(b\) instantiates a strict law’\(^{229}\). This is so because in saying that \(a\) instantiates a strict law only under a description, we make the occurrence of the term ‘\(a\)’ in the sentence ‘\(a\) instantiates a law’ intensional. And so even if \(a = b\), what is claimed as true about \(a\) might well be false about \(b\). Another thing that we have to remember is that causation is a relation between events as such and not

\(^{228}\) ‘Mental Events’, p.208.

events under a certain description. Consequently if \( a \) is a cause of a physical event and so is \( b \), then they are one and the same event. All that \( a \), being the event under the mentalistic description, fails to do is be subsumed under a law. But the event (as such) having a physical description can be subsumed under a strict law and so the event (as such) becomes the cause. And so when Davidson says that there are no psychophysical laws, what he means is that there are no strict laws featuring both mental predicates and physical predicates. May be it would not be incorrect if we looked at this in the following way. From a metaphysical point of view the fact that two events are causally connected does not depend upon how they are described. But in order for us to find out if there is such a causal relation between the two events (i.e. from an epistemological point of view), we have to find out if there can be some descriptions of the events under which they can be subsumed under a strict law. This is perhaps how we can solve the apparent inconsistency in Davidson's claims.

In this way Davidson manages to hold on to all but one presupposition of the physicalist. He believes in mental causation and manages to show how mental causation is possible. He does not have to believe in over-determination. And he also holds on to the completeness of physics assumption. The completeness of physics assumption is perhaps not so evident in the case of anomalous monism. But if we take a closer look into Davidson's argument, then we shall see that it follows from two of Davidson's assumption. On the one hand, Davidson believes in the nomological character of causation. And so if there is a causal connection between two events, then there is a description of them under which they can be subsumed under a strict law. The other thing that Davidson is saying is that there are no psychophysical laws, as
the events under the mentalistic description cannot be subsumed under a strict law. But in case they are indeed causally related, then there would be a physical description under which they can be subsumed under a law. Consequently, the mental causes (events under mentalistic description) would not suffice in fixing the physical effect, and since the nomological connection holds only between physical events (events under physical descriptions), it is the physical causes that determine the physical effects exhaustively. And in this way you come up with the completeness of physics assumption.

For Davidson the problem arises because physicalists have adhered to another presupposition, viz. that causes have their effects in terms of some properties. We shall now discuss what is involved in giving up this assumption. It seems a bit bizarre to give up this assumption, as this usually is the way in which we understand causation: if X causes Y, then it does so by possessing some property by virtue of which it becomes the cause of Y. Davidson has questioned this very way of thinking about causality. Davidson does not believe that mental states/events become causally efficacious by possessing properties, in other words the causal efficacy of mental states should be understood in terms of properties. But as we have seen from the discussion of various kinds of physicalism, the problem of mental causation is often seen as a problem addressing this question: Do mental events have effects in terms of their mental properties? This question is again brought against Davidson's anomalous monism by the critics, but what the critics forget is that Davidson has questioned this very way of looking into the problem.
Some critics say that the way that Davidson construes his thesis is incoherent with his claim that causes have their effects not in terms of any properties. These critics have brought this charge mainly because they think that in order to make sense of Davidson's anomalous monism, we have to understand causality in terms of property. Though I shall show that they are wrong in assuming this, yet let me first say why they might at all come up with their criticism. They feel that Davidson's anomalous monism seems to suggest that he adheres to the view that events are causally efficacious in terms of some of their properties. 'To say that two things are not in law-like connection under certain descriptions is to say that certain of their properties are not in law-like connection, or perhaps, that the things are not in law-like connection in virtue of their properties.'

231 From the critics' reading of the Davidsonian thesis it seems that Davidson construes the notion of cause in terms of properties, and it is because of this that he is saying that there is no law-like connection between the mental and the physical. And it is because of this that Davidson is saying that the mental events are actually identical with the physical events. But if mental events are indeed physical events, then we have to deny the claim that there are no psychophysical laws. Here we have to distinguish between two claims: the claim that the mental as mental causes the physical, and the claim that the mental as physical causes the physical. Or in other words, the mental can cause something because it is mental and the mental can cause something because of being something else (in this case, in some sense physical). In the second case, it is not the mental which is doing the causing. And so the question arises as to which of the two is Davidson subscribing to. And can he answer the (apparent) contradiction that his thesis is facing?

230 See his 'Causal Relations', in Essays on Actions and Events.
The problem is this. As events are particulars, they are regarded to have properties. And so we can ask about an event if it was causally efficacious by virtue of its mental properties or its physical properties. If it was the first case, then there should be strict laws mentioning the mental properties of the event in terms of which the effect can be predicted or explained. This would obviously go against Davidson’s view of Anomalous Monism. On the other hand, if we accept the second, i.e. it is in terms of its physical properties that the event manages to be causally efficacious, then mental properties would become causally totally inert and epiphenomenal. It can be seen in a form of dilemma. If the causal efficacy of the mental events is understood in terms of their mental properties, then Davidson’s view that there are strict laws connecting events which are causally related has to be given up. On the other hand, if the causal efficacy of the mental is understood in terms of the physical properties of the event, then the mental becomes epiphenomenal.

It should be clear from our discussion of Davidson’s thesis that one cannot read Davidson in this way. We will discuss in detail why this way of reading Davidson is (to say the least) not quite correct.

4.6 Davidson on the Causal Efficacy of the Mental

The heart of Davidson’s response to this is that we should keep in mind the difference between Causation and Causal Explanation. And we have to remember that under the Davidsonian scheme of things, there is no issue about whether or not mental properties are causally efficacious, as only events can be efficacious. There is however an issue about
whether or not mental properties (if you have to have properties at all) are causally explanatory. The critics of Davidson which we have so far discussed have not kept this in mind.

To start with, we have to remember that Davidson would not speak in terms of properties and so would not speak in terms of mental properties either. And even if we allow ourselves to talk in terms of properties, Davidson could say this: If two events are causally related, then there would be a strict law linking them, but because of the anomalism of the mental, this strict law will not mention the mental properties of the event. But this does not mean that these properties are irrelevant to causal explanations. We can never say that they are causally completely irrelevant, because they figure in non-strict laws which we appeal to all the time in giving causal explanations of those events. For Davidson that the mental is causally efficacious is a matter regarding whether or not the mental properties do figure in causal explanations. (This might seem to be whole of the Davidsonian support to the thesis regarding the causal efficacy of the mental. And hence it might seem pointless to go into any other kind of support from Davidson. One could say that all that we need to do is to articulate this. But it seems to me that maybe there is more to be said. And that is why we find people like Sosa, Kim and Fodor pressing Davidson to come up with a more full-blooded explanation of how the mental is causally efficacious. And that is why we find Davidson bringing in supervenience as a support of his view about the causal efficacy of the mental.)

To start with, I think since Davidson places so much importance on what figures in causal relation, he needs to justify why mental predicates (properties, if we allow ourselves to be a bit sloppy in our
terminology) get mentioned in causal explanations. And I feel that this is why he brings up supervenience.

Let us see more in depth how Davidson would defend his theory against the charge of being epiphenomenalistic. Anomalous monism is a view that mental entities (objects and events) are physical entities, but mental concepts are not reducible to physical concepts. Anomalous monism (or AM from now on) is derived from three premises (or P from now on):
1. Mental events are causally related with physical events.
2. Singular causal relations are backed by strict laws.
3. There are no strict psychophysical laws.

The mains objections that Davidson faces are that the arguments for not accepting any psychophysical laws are obscure and the premises are inconsistent and that AM makes mental causation inert. Davidson tries to show that the premises can be held together and that AM is a tenable thesis. What is most important for Davidson is to show that AM+P doesn’t make the mental causally inert.

As we have already said, for Davidson events are particular entities which form a part of our common-sense ontology. And if they are entities, then like any other entities they can be described in very many different ways, and these descriptions can be logically independent of each other. The causal relation is a relation between events. Such a relation exists even if there was no one to describe them. Of course if mental properties/predicates were completely independent of physical predicates/properties, then it would be difficult to find some causal relation between the two kinds. Davidson says that they are not
independent, but what we need to do is to get the relationship, between the two kind of predicates/properties, right. And it is here that the relation of supervenience comes in.

Historically speaking, philosophers of mind started speaking of a supervenience relation between the mental and the physical in order to make a non-reductive materialist thesis acceptable. And Davidson surely wanted to achieve that. In addition to this, as we shall soon see, he thinks that the supervenience relation would help him to show how the mental is causally efficacious. However, first let us see whether this relation helps Davidson to establish his non-reductive materialism. We need to show that he indeed does manage to establish his non-reductivist thesis. This is needed because critics like Kim have tried to show that the kind of supervenience Davidson is opting for neither helps him to establish his physicalism, nor does it help him to show how the mental is causally efficacious. Kim's main objection is that a non-reductive physicalism is no physicalism worth the name and even if it is, such a physicalism would actually make the mental epiphenomenal.

Let us first see what this relation is. The relation originally was construed in terms of properties and it was taken to be this: A property is supervenient on another class of properties in case the supervenient property doesn't make any distinction among entities which the subvenient property doesn't. It couldn't be the case that two events which had exactly the same physical properties had different mental properties. Every difference in mental property is accompanied with some change in the physical property. We have to remember that Davidson formulates the relation in terms of predicates rather than
properties. And so it would be better to conceive the relation in terms of predicates. A predicate \( p \) is supervenient on a set of predicates \( S \) if and only if \( p \) does not distinguish any entities that cannot be distinguished by \( S \).

Let us now see how the relation of supervenience helps Davidson to prove his non-reductive physicalism. To start with, the thesis of anomalism of the mental itself shows that there can be no reduction of the mental to the physical. A part of the anomalist thesis says that there can be no physical predicate (no matter how complex it may be) to which a mental predicate can be reduced. What is meant by this thesis is that we cannot find bridge laws linking the physical predicates to the mental ones. Usually we speak of bridge laws in cases of theories which are reducible to each other. We can have bridge laws between predicates (or properties as some would prefer) which are necessarily co-extensive. Such laws do not link the physical predicates with the mental ones. And hence there can be no reduction of the mental to the physical.

Traditionally a non-reductive thesis was thought to indicate that the mental predicates should be eliminated. Davidson obviously does not want to opt for an eliminativist thesis about the mental. But does this mean that Davidson is opting for a dualism of some sort? Definitely not. He finds out a close relation between the mental and the physical. Strong enough to link them causally and weak enough to resist reduction. And such a relation is the relation of supervenience. The relation of supervenience ensures a sort of relation of dependence between the mental and the physical, but resists reduction. The idea of supervenience held for the philosophers of mind a third option in between

\[ \text{[232] See his 'the Myth of Non-reductive Materialism', in his } \textit{Supervenience and Mind}. \]
eliminativism and dualism: the option of non-reductive physicalism. Now what is this notion of supervenience. The key ideas of this notion of supervenience are: (i) covariation of predicates/properties, (ii) dependence, and (iii) non-reducibility. These capture, for Davidson, the right kind of relation between the physical and the mental. This is so because Davidson thinks that we can have no variation in mental predicates without any variation in physical predicates. He also says that this is an indication of how the mental depends on the physical. This is a one-way dependence: from the mental to the physical and not the other way around. If the relation was anything more, then there would be the risk of reduction. Moreover, that the mental cannot be reduced to the physical is established by Davidson with the help of the thesis that there can be no bridge laws linking the mental and the physical predicates.

Kim says that the relation of supervenience can be of different kinds depending upon the strength of the relation of dependence between the supervenient and the subvenient. And he thinks that Davidson's supervenience can be called weak supervenience. This is the definition of weak supervenience that Kim offers:

'A weakly supervenes on B if and only if necessary for any \( x \) and \( y \) if \( x \) and \( y \) shares all properties in \( B \) then \( x \) and \( y \) share all properties in \( A \) - that is, indiscernability with respect to \( B \) entails indiscernability with respect to \( A \).

We shall call \( A \) the supervenient family and \( B \) the supervenient base (family); properties in \( A \) are supervenient properties, and those in \( B \) are base properties.'

---

233 We shall discuss this in detail when we speak of how the mental becomes causally efficacious.

234 Kim, 'Concepts of Supervenience' in Supervenience and Mind, p. 58.
What this means, according to Kim, is 'Necessarily (that is in every possible world), if any x and y (in the domain) are indiscernible in P ('P-indiscernible' for short), x and y are M-indiscernible'. This shows that two entities within a world cannot vary in terms of their mental predicates/properties without varying in their physical predicates/properties. Kim thinks that this is not a strong enough relation to ensure the kind of dependence of the mental on the physical that is required for physicalism, as under physicalism if two worlds are physically indistinguishable, they should be indistinguishable in mentalistic terms as well. But this might not happen under weak supervenience. And Kim also feels that this type of dependence would not ensure causal relation between the two. Kim thinks that it will be difficult for Davidson to combine the two ideas of non-reduction and dependence together in one relation. Kim says the main difficulty has been this: if a relation is weak enough to be non-reductive, it tends to be too weak to serve as a dependence relation; conversely, when a relation is strong enough to give us dependence, it tends to be too strong — strong enough to imply reducibility. Is it possible for Davidson to answer this charge brought by Kim? Let us see.

First of all we know that Davidson would never opt for a relation between the mental and the physical which would be so strong as to entail a reduction between the due. And we know why he doesn’t. Secondly he feels that his weak notion would suffice to show how the mental is causally efficacious. In order to understand this claim we have to see what is essentially wrong in seeing the causal relation in terms of properties.

235 Kim on Supervenience in Guttenplan (ed.) A Companion to the Philosophy of Mind.
Davidson says that given the conception of events and causality there is no sense in saying about an event that it is a cause 'as' anything at all. We say X is the cause of Y simpliciter, not as this or that. Davidson disallows this locution in the realm of causality. But there is a realm where we talk about properties (mental predicates) and it is the realm of causal explanations. And it is here that the mentalistic predicates have their place, not in strict laws of physics but in non-strict laws which we take recourse to when we give causal explanations. But not all things get mentioned in causal explanations. Even if they do, there is a reason behind this. And in this case it is the presence of the supervenience relation between the mental and the physical which is the reason for which the mental gets mentioned in the explanation of the physical.

Maybe it is not correct to think that Kim and other critics are charging him of inconsistency. What is more important for them is this: the truth of the first premise of anomalous monism does not ensure the causal efficacy of the mental. Now is it wrong on the part of the critics to think that AM+P is a form of epiphenomenalism? Kim says in 'Can Supervenience and "Non-Strict Laws" Save Anomalous Monism?', that the critics have sometimes failed to make a distinction between two claims: (i) AM+P entails the causal inertness of mental properties. (ii) AM+P fails to provide mental properties with a causal role. Davidson would definitely resist (i), and it is certainly something that cannot be said about Davidson. But how would Davidson answer the second charge? Kim says that if a theory says a lot about the causal power of the physical properties and say nothing about the causal powers of the mental ones, then certainly that theory is epiphenomenal with regard to the mental properties. Kim sees Davidson's mention of the
supervenience claim and also the mention of non-strict laws in addition to AM as an attempt to restore the causal efficacy of the mental. This, for Kim, is actually to tacitly acknowledge that within the framework of AM+P one could not accommodate the causal efficacy of the mental properties.

One important mistake that Davidson thinks that the critics are making is that they forget that causal relation, for him, is an extensional relation between events, and if we say X is a cause of Y as f, then it would no longer be a dyadic relation. If causality is a relation between events, then it holds no matter how the events are described. It is a completely different thing to say that they instantiate a law if described this way or that. Two events if, they are causally connected, would be so connected no matter how they are described. All that AM was designed to show was that 'there can be descriptions of two events (physical descriptions) which allow us to deduce from a law that if the first event occurred the second would occur, and the other descriptions (mental descriptions) of the same events which invite no such inference'. We have tried to explain this point earlier in Section 4.5.

Now is it right to think that the critics are making a mistake here? Have the critics really turned the binary relation between cause and effect into a multi-term non-extensional relation? Brian McLaughlin, in his paper 'Davidson's Response to Epiphenomenalism', says that Davidson's claim that the critics are turning a dyadic relation between two events into a four-termed relation is not true. In saying that X and Y differ in their weight, we specify in virtue of what does X and Y differs,

---

238 P 32.
viz., their weight. But this doesn't mean that we are admitting an extra entity as weight other than the weight of X and the weight of Y. But Davidson would say that the need for describing the events this way or that only arises in the case of causal explanation, as it is linked with the specific interest with which we embark upon giving an explanation. But maybe the critics are saying something else. As Kim points out, what the critics are saying is that the extensional binary relation is not enough, as we need a way of talking about the causal roles of mental properties. Kim wants to say that what the critics have said is perfectly consistent with the thesis that causality is a binary extensional relation between events. But still the critics feel that in order to state what actually constitutes the relationship, we have to talk about the roles of properties in this causal relation. Kim says that it makes sense to ask the question 'What is it about events c and e that makes it the case that c is a cause of e?' And Kim thinks that a satisfactory answer to this would be of the kind 'This is so because c is an event of the kind F and e is an event of the kind G'.

Kim goes on to say that maybe AM+P+Davidsonian Supervenience does not provide a satisfactory account of causal relevance of mental properties. It is true that there is a distinction between causal relevance and causal efficacy. I am mentioning this difference as some philosophers, to my mind, are making this distinction, for instance, Kim. It is as though properties/predicates that get mentioned in causal explanations are causally relevant, and those which form a part of a strict law are the ones which are causally efficacious. I do not believe in this distinction and nor does Davidson. But critics have mistakenly assumed that he is making such a distinction

239 In 'Can Supervenience and "Non-Strict Laws" Save Anomalous Monism?'. 
or at least if one has to make sense of what Davidson is saying one has to invoke this distinction. Anyway assuming that the distinction exists, one could say that even the epiphenomenalists would agree that mental properties are causally relevant. And if Davidson is saying that since the mental predicates get mentioned in the non-strict laws, they are causally relevant (though he confuses and uses the word 'efficacious' in this case according the critics), then all that supervenience (if it is not automatically linked with psychophysical laws) can show is that mental properties are causally relevant. But does that mean that they are causally efficacious too? It seems that mere causal relevance cannot support the causal-explanatory "because", which we need in rationalising explanation. Now if this be so, then what would happen to AM+P+the existence of Non-Strict Laws? Kim thinks that there are serious difficulties in admitting non-strict laws along with AM+P. Much of it depends upon what we regard as a non-strict law ('or what the much-bandied phrase 'ceteris paribus' means when it qualifies a law?') Kim says that it seems that where there is a non-strict psychophysical law, there is a strict psychophysical law waiting to be discovered. Do non-strict laws have no nomological force at all? If they do, then anomalous monism cannot allow for them either. There may be other reasons for being wary of non-strict laws, if you are an anomalous monist. What is the causal role of mental properties which get mentioned in the non-strict laws? If it is the physical properties which get mentioned in the strict laws (and hence it is those properties alone which are causally efficacious), then what causal role can mental properties have? But aren't the critics putting their own words in Davidson's mouth? Davidson knows what he means when he speaks of causal efficacy. He is not speaking of causal relevance. He is indeed speaking of causal efficacy. The only point that he wants to make is that we should
keep the levels of explanation and causation apart. And it is on the level of causal explanation that the question of causal efficacy arises. And it is this point that the critics fail to understand.

On the other hand, in answer to Kim, he would say supervenience does entail that any change in the mental property \( p \) of an event \( e \) would be accompanied by the change in the physical properties of \( e \). But this does not mean that a change in \( p \) in other events, would be accompanied by the change in the physical properties of those events. If supervenience would entail that, then it would conflict with AM\(+\)P. There is no entailment from supervenience to the fact that the same physical properties change with the same mental properties. Davidson thinks that Kim is of the opinion that Davidson is opting for a version of supervenience which would entail that if we were to remove all the mental properties of the events of the world, then we would have no consequence on how physical properties are distributed over these events. But this is not what Davidson wants to say. What Davidson wants to say is that if we consider two events with the same physical properties, but one with a mental property and the other without that mental property, then these events cannot be the same event at any rate.

May be the lurking difficulty that the critics find in making sense of Davidson is because they are mistakenly thinking that when Davidson talks about the causal efficacy of mental events he has to say that it is the mental properties which are efficacious. However that is not what he wants to say. For Davidson, it is the events which are causally efficacious and not the properties. This he thinks because properties are abstract and they cannot be causes of anything. Causal relations are always between events and not between properties. But of course when we come to state
the laws, then we have to classify events in one way or the other. Loose, non-strict causal laws connecting mental events with physical events which are of great practical value. But they record regularities and they are not strict laws. Davidson claims that people are confusing strict laws with the multitude of possible explanations of relations between events.

There is however something that can be said on behalf of the critics. Suppose the critics are not bringing the charge of inconsistency against Davidson and suppose that the critics do recognise the difference between causation and causal explanation. Still one could go on to say that if we accept that causal relations are objective, then would there not be some explanation which would be better than all others as it picks out the real causal features of the world?

Davidson would definitely resist the idea of there being one true explanation. The merit of an explanation, for him, depends entirely upon the interest that motivated one to look for an explanation. And if one presses him to show why at all would the mental figure in any kind of explanation, he has supervenience to support it.

In understanding Davidson's view we have to keep in mind the two main planks of Davidson's view about causation, viz. that causation is a relation between particular events; and that there is a serious difference between causation and causal explanation and his reservations against properties.
4.6 Anomalous Monism and Externalism

That Davidson is an Externalist and that his anomalous monism is in perfect harmony with this externalism is clear from his 'Knowing One's Own Mind'. In connection with the question of whether first person authority can be fitted into an externalistic thesis, Davidson says that it can be done. He says that mental states are "inner" in the sense of being identical with states of the body, and so identifiable without reference of objects or events outside the body; they are at the same time 'non-individualistic' in the sense that they can be, and usually are, identified in part by their causal relations to events and objects outside the subject whose states they are.240 This clearly shows that he is an externalist. His externalism is well in tune with his views on mental causation. He clearly says that externalism can well blend with identity theories but only of the token identity variety and not type identity variety. And he also claims that it is well in tune with his specific kind of token identity theory, viz. anomalous monism. He says that there is a tendency amongst philosophers to think that externalism leads to a rejection of physicalism. But he says that it does no such thing. He says that if two mental states differ in their content (externalistically conceived), then they are indeed different. For instance Ted and Ed's water/ was water thoughts are indeed different 'even in the physical world; their causal histories are different241. Davidson says that it would be wrong to conclude from the mere fact that some mental states and events are individuated in terms of the relations they have with the outside world, that there can be no identity between the mental and the physical.

240 Davidson, 'Knowing One's Own Mind', p. 48.
241 'Knowing One's Own Mind', p. 59.
All that externalism about the mental does show, according to Davidson, is that type-type identity theories are implausible. Davidson thinks that the main reason that other philosophers have for saying that externalism cannot blend with identity theories is this: These philosophers wrongly claimed "that it is "absurd" to think two people could be physically identical... and yet differ in their ordinary psychological states". Davidson claims that his anomalous monism can show that 'people can be in all relevant physical respects identical while differing psychologically'.

So we see that Davidsonian externalism can blend well with his anomalous monism. This anomalous monism has a good answer to the general problem of non-physical causation. Moreover Davidson does not face the specific problem of content causation that externalists were thought to face, in spite of the fact that Davidson is indeed an externalist. This is because if one gives up the conception of causation which says that events are causally efficacious only in virtue of something, then she would manage to show that this specific problem of content causation will dissolve. Since causes produce their effects all by themselves and not in terms of anything, the question whether mental states manage to cause physical states in terms of their external content or in terms of their internal content does not arise any more.

---

242 This point has been discussed in Chapter One.
243 'Knowing One's Own Mind', p.60.
Conclusion:

In rounding up let us see what this thesis has tried to achieve and what could be the further queries that it might lead us to. This thesis has aimed at making sense of the externalist-internalist debate concerning how we can individuate mental states. The importance of the question about how mental states can be individuated has been recognised by all. Of course one needs to have a criterion of individuation for the entity about which one wants to investigate and theorise. And so before we even start thinking about the mind and the mental we need to know what the mind and the mental are like. The externalist-internalist debate centres around this crucial question. Historically, this debate has a very special place in philosophy. I feel that this debate has set the stage for a real dialogue between philosophy of mind, epistemology and philosophy of language. The debate started off with Putnam's theories of reference and hence we could say that philosophy of language is its mother. We have seen in the thesis how metaphysical questions about the mind are inextricably linked with semantic and epistemic issues. Though the debate is a debate in the philosophy of mind, it has serious ramifications in philosophy of language and epistemology. The questions about the mental and about reference and about first person authority are intermingled in such a way in the debate, that there is no way of answering any one of these kinds of questions, without taking into account what could be the answers to the other two kinds of questions. The interaction of mind, language and knowledge makes this debate at the same time attractive and difficult to decipher. I have tried to bring out this dynamic character of the debate and the mind-boggling
interplay of (metaphysical, epistemological and semantically) intuitions involved in it.

I have not tried to come up with a solution of the debate (nor a resolution of it). I feel that there is no way of understanding the debate without looking into the intuitions that have given birth to it and the intuitions that still sustain it. This thesis is an enquiry into the possible internalist intuitions, and hence this critic of the possible presuppositions/assumptions of the internalists. I am not suggesting at any point that externalists do not have any presuppositions. They do, and those are equally open to scrutiny and criticisms. But I have not aimed at doing that in this thesis. A natural extension of this work would, of course, be a research into the possible externalist intuitions. However, some of the intuitions that lead philosophers towards externalism have already emerged in this critic of the internalist intuitions. I shall try to say, once again, what I think are the internalist intuitions which have made the debate take its present shape. And I would like to hint at the possible externalist intuitions that can be extracted from the discussion of the debate here.

But before doing this I have one important work to do: One might question the importance that I assign to this debate. Those, who unlike me, do not find this debate crucial to the serious concerns in the philosophy of mind could actually have a point. And I would first like to address that. As I have said, right in the beginning of the first chapter, I see this debate as a debate especially about intentional mental states. If this is the case, then one might find it difficult to see how this debate can throw any light on how mental states, in general, can be identified. However, this is not quite right. There is a lot to be learnt form the
debate. This is so because the internalists have tried to point out this very point (i.e. the importance of having a way of individuating mental states in general), to the externalists. And in doing so, the internalists have come up with a view about individuation of mental states, which are applicable to mental states in general. In this respect the contribution of the internalists in the debate is most remarkable. I think that the externalists acknowledge this, and hence we now find more moderate externalistic views.

As we know that the debate between the two parties is regarding mental states, the very existence of which entails the existence of the objects towards which they are directed. A possible way of handling the impasse we face, without avoiding the metaphysical issue involved here, is by saying that there are two kinds of mental states, those which are directed towards existent objects and those which are directed towards non-existent objects. And so we need two different kinds of criterion of individuation for them. This approach to the problem is motivated by the thought that it is perhaps wrong to look for an invariant individuation story for all kinds of mental states. But there seems to be a sort of general metaphysical argument in favour of opting for a uniform individuation story. If we keep in mind why we need to have principles of individuation, then this will be clear. Things are put into different metaphysical categories according to their principles of individuation. Now if we have two distinct principles of individuation for the same metaphysical category, then that would pose a threat regarding whether or not the category in question is a genuine metaphysical category. This
would be happening if we opt for two distinct principles of individuation.\footnote{This was pointed out to me by P. K. Sen.}

There seems to be good reasons for supposing that we need just one individuating principle for all kinds of mental states bringing out the basic common feature or characteristics of those mental states. I want to make a linguistic point here. It seems that the externalists are overly impressed by the behaviour of factive verbs so often used in the description of mental states, or in other words in making mental predication. A typical example of such a factive verb would be ‘know’ or ‘see’. Let us concentrate upon the second. If I say ‘I see an approaching bull’, I do not just say in what mental state I am, I also say what is there in the external world. Not to make too fine a point of it, what I say is something like the following conjunction: I am having an approaching bullish visual experience and there is (in fact) an approaching bull. This conjunction would certainly be false if the second conjunct was. But it is equally certain that the falsity of the second conjunct does not entail the falsity of the first. There may not be in fact any approaching bull and yet I may have the experience I have.

The claim which I make in saying ‘I see an approaching bull’ may be construed as a claim to the effect that I am in the mental state of seeing an approaching bull. Once this way of making the claim is permitted, one may demand to know whether I am really in that state. Isn’t it obvious now that I must have some way of identifying or individuating an arbitrary mental state as the state in which I claim myself to be in? This immediately leads to an externalist construal of the criterion of individuation: if seeing an approaching bull entails the
existence of an approaching bull, then how can we individuate a mental state as a state of seeing an approaching bull without looking at my environment. What the externalists are missing is that it is not the case that mental states fall into two distinct categories calling for two distinct ways of individuating them, but the verbs used for making mental predication do fall into two different categories, some factive and some non-factive. Whether or not the verb of mental predication is factive or non-factive, there is a mental state which it is used to predicate. This mental state itself does not incorporate the object as something intrinsic to it, as something in terms of which it should be individuated. Considering this it seems that there must be a way of establishing the first component of the conjunct irrespective of the second. This independence is brought out by the internalists when they speak about empty mental states and the empty behaviours that they give birth to. There are two internalist intuitions nested here. The first is that we need something invariant amongst all the mental states and it is only what it is like for the subject to be in the mental state which is common here. The second is that in order to determine this common character we do not need to look outside into the environment. And the internalists point out that this is evident from the fact that we do not need to look at the external world in order to explain the empty mental states and empty behaviours. This has led internalists to speak about phenomenal characters. The question regarding the importance of the phenomenal over the intentional has also been central to this debate. And we have seen how views of the internalists and the externalists differ radically about the phenomenal aspect of mental states.

In spite of the fact that we might appreciate the internalists view that there is a need for having a way of specifying the subjective aspect
of mental states without the need for having to mention the world-involving aspect of them, still we can ask at least one question: Is there indeed a way of establishing the one without having to do anything with the other? If the internalists want to make use of this intuition to the benefit of their own thesis, then they have to conceive of the common element in a particular way. They have to take this common element to be something which is subjective *par excellence*. And we have seen in the second chapter the problems that there are there with such a thoroughly subjective notion. I would like to express one simple externalist intuition that goes against such a purely subjective notion. Let us see what the internalists are saying. They are saying that the subjective element in mental states is what makes the states *mental*, and hence it is more reasonable to individuate mental states in terms of this element. I would like to ask the internalists just one question: If our mental states were so subjective, then how is it possible that we manage to convey to others what we have in our minds? How is it that we understand what others have in their minds and others understand what we have in our minds? Certainly this indicates that we must be sharing something, in terms of which we can make sense of the communication that takes place amongst us. It couldn't be the subjective element which is allegedly invariant to all our mental states which provides the common ground for our communication. If we ask ourselves, 'What is it that we share with others?', the answer would be, 'We share the world'. And so in order that others understand what I have in mind the world has to be reflected somehow in the mind. There must be some essential world-involving part of our mental state which provides the common ground in which communication takes place. This externalist intuition indicates that there must be an essential mention of the external in the individuation of the mental. This intuition respects the internalist intuition that we need to
find a common invariant element in all mental states, as this view does not speak of two different kinds of individuations for empty and non-empty mental states.

But of course there are other worries with this external world-involving element, as is demonstrated by all the internalist arguments against the externalists. Of course the main worry would be centring around the question as to how we can make sense of the external world-involving element of empty mental states and the empty behaviours they elicit. In response to this the externalist might try to show that the understanding of these deviant empty cases is somehow parasitic upon the understanding of the non-empty cases. This we have found in the second chapter. The other major worry that this externalist way of individuation of mental states gives rise to is that this seems to deny first person authority. The internalists, as we have seen in the first two chapters, adhere to the Cartesian insight that the subject is usually right about her own mental states. The emphasis on the importance of finding a common element invariant to all mental states, we have seen, goes hand in hand with Cartesianism to form a support for the narrow individuation of mental states. And these two together make it possible for the internalists to explain empty thoughts and empty behaviours. I would like to keep this set of intuitions together. We have seen, from our discussion on the possible externalist responses to the charge that externalism is incompatible with the idea of first person authority, that this charge can be met if we opt for a slightly wider notion of what is intrinsic to us. The internalists suppose that in order to account for the immediacy with which we know our mental states we have to take that element of mental states as individuative, which is strictly intrinsic to the subject. This is so because it is believed that it is only about what is
intrinsic to us that we can have the kind of immediate access that is required in self-knowledge. The idea of the intrinsic has been central to the debate, as the two contending parties have two different kinds of notions of the intrinsic. The internalists take the strictly internal to be intrinsic while externalists might take a more relaxed stand on what could count as intrinsic. We have discussed this again in the third and the fourth chapters. (We have, however, discussed this in the context of another set of internalist intuitions, intuitions which we will talk about shortly.) I have tried to show how all these together are responsible for the first presupposition that I think that the internalists are adopting: We can consider two mental states to be the same (i.e. type-identical), if the subject cannot discriminate them.

There is however another set of internalist intuitions (which are not totally detached from the previous set) which I think have given rise to the second assumption, viz.: We can consider two mental states to be the same if they give rise to the same behaviour, *ceteris paribus*. In asking the metaphysical question about the individuation of mental states we might be influenced by another kind of consideration. We might ask ourselves: Why do we at all need to talk about mental states? The answer would be that mental states are believed to be things that move us to act. And so we need to speak about mental states, because they help us to explain our actions/behaviours. This is an intuition which the internalists and the externalists share. Associated with this intuition, is the thought that the evaluation of the worth of a theory about the mind is tied with the question of how best could the role of the mental states in the explanation of behaviour be captured in that theory.
It was commonly believed that mental states come into the explanations of behaviours as they are the causes of behaviours. And this made philosophers think that we need to individuate mental states in terms of their causal roles, or causal powers. On this issue we found internalists and externalists to have different things to say. We find them differing even in their conception of behaviour, and consequently differing about what one should regard as the explanandum of psychology. And it is here that we find differing intuitions amongst the two debating views. We find them actually differing about the very nature of psychology itself, which in turn leads to their difference regarding what should be regarded as the explanandum of psychology. And so we find internalists and externalists differing in their views about mental states, by virtue of differing in their views about the explanandum of psychology (i.e. behaviour) and psychology itself. Related with this issue is the question what kind of psychological theory we are looking for: Folk Psychology or Scientific Psychology. And so the debate between folk and scientific psychology has much to say in the internalist-externalist debate as well. The desire to fit the mind into a neat structure of a science has led philosophers towards internalism. In such a science we individuate entities in terms of their causal powers and it seems that it is impossible for us to reconcile causation with content, if we take psychological states to be wide.

Along with this came another strong intuition of the internalists: the belief in the principle of psychophysical supervenience. Internalists, as we have seen, believe that causal powers supervene on 'local microstructures' or internal properties. They also believe that it would not be possible to have a science of the mind if we did not respect this supervenience. The twin earth thought experiments and the way that
they have been used by the externalists to their own benefit deny this supervenience. And that is why internalists like Fodor have rejected the wide content theory. These internalists feel that since under the externalist thesis mental states are relationally individuated, they cannot supervene on brain states which have to be non-relationally individuated. And they also say that the reasons for which one takes brain states as non-relationally individuated and individuable are the reasons which are there even for the mental states. Scientific theories are usually taken to be internalistic. This idea is closely related to the idea of causal power. In scientific psychology states are individuated in terms of their causal powers and it is only when states differ in their causal powers that they are taken to be different. There are quite a few things that can be said about this. This internalist view has been motivated by a particular notion of behaviour which is narrow. We have seen in the third chapter what are the problems with such a narrow notion of behaviour. There is another crucial intuition that has gone into this internalist view. It is the view that if anything could possibly cause behaviour, it has to do so only by mediation of intrinsic properties of the subject. And since external objects and the external environment cannot form the intrinsic properties of mental states, we cannot say that mental states become causes in terms of their wide content. This is the content causation problem which we have spoken of in the fourth chapter. Here again the question regarding what could count as the intrinsic to a mental state comes in. And we have seen that internalists and externalists have divergent intuitions about this.

We have seen how all this has given birth to the mental causation problem in a new way. However we have seen how, following the Davidsonian line, we could demonstrate that the mental causation
problem does not have such a close relation with the content causation problem and hence the internalist-externalist debate. The solution that is suggested however relies heavily on our conception of the causal relation. This is a hard-core metaphysical issue and our views, intuitions about this will shape our responses to the mental causation problem, as well as the content causation problem.

All these intuitions have made the internalist-externalist debate what it is. And I feel that there is no way of understanding the debate without a careful scrutiny of the conflicting intuitions that sustain it. The lack of understanding of the intuitions behind the two contrasting theses has aggravated the confusion and resulted in un-resolvable disputes amongst philosophers. To solve or dissolve the debate, we have to question the diverse intuitions that have made the confusion about the debate more confounded. This thesis has been a scrutiny of the intuitions of one of the conflicting sides, it has been On Internalist Intuitions.
Bibliography

Anscombe, G.E.M.


Armstrong, D.M.

*Belief, Truth and Knowledge*, Cambridge, 1973

Bach, K.

'De Re Belief and Methodological Solipsism', in Woodfield (ed.),


Blackburn, S.


Block, N.

*Readings in the Philosophy of Psychology*, 2 vols, Harvard University


'Qualia', in Guttenplan (ed.), *A Companion to the Philosophy of


'Can the Mind Change the World?' in G. Boolos (ed.), *Meaning and

Method: Essays in Honor of Hilary Putnam*

Bilgrami, A.

'Can Externalism be Reconciled with Self-Knowledge' in Pessin,

Goldberg (eds.), *The Twin Earth Chronicles*, M.E. Sharpe, Armonk,

New York, 1996

Blackburn, S.


Boer, S. and Lycan, W.

Boghossian, P.


Broad, C. D.


Burge, T.


'Individualism and Self Knowledge' in Cassam (ed.) *Self Knowledge*.

Butterfield, J. (ed.)


Cassam, Q. (ed.)


Chalmers, D.


Chatterjee, A.

'The Intentional and The Phenomenal', paper presented at seminar on 'Consciousness and the Phenomenal Mind', Lucknow, India, February, 1999

Crane, T.


Crimmins, M.


Davidson, D.


'Causal Relations', in Essays on Actions and Events.

'Mental Events', in Essays on Actions and Events 'Psychology as Philosophy', in Essays on Actions and Events.


'Knowing One's Own Mind', in Quassim Cassam (ed.), *Self-Knowledge*

Dennett, D.


Donnellan, K.


'Proper Names and Identifying Description', *Synthese* 21, 1970.

Dretske, F.


Dummett, M.


Evans, G.


Fodor, J.


Geach, P.T.


Gibbons, J.


Guttenplan, S. (ed.)


Harman, G.


Heal, J.


Hilton, J.M.


Heil & Mele (eds.)


Hintikka & Vermazen (eds.)

Honderich, T.

Hopkins, J.

Hornsby, J.
'Physicalist Thinking and the Conception of Behaviour', in Pettit & McDowell (ed.), Subject Thought and Context.

Humberstone, I. L.
'Intrinsic/Extrinsic' Synthese 108, 1996

Jackson, F.
'What Mary Didn't Know', in Journal of Philosophy, 83.

Kaplan, D.

Kenny, A.

Kim, J.
'Psychophysical Supervenience', Philosophical Studies, 1982.
'Can Supervenience and "Non-Strict Laws" Save Anomalous Monism?', in Heil and Mele (ed.), Mental Causation.
'Events as property exemplifications' in Supervenience and Mind, Cambridge University Press, Cambridge, 1993
'The Myth of Non-reductive Materialism', in his Supervenience and Mind
'Concepts of Supervenience' in Supervenience and Mind,
Klein, M.


Kripke, S.


LePore and McLaughlin (eds.)


Lewis, D.


Loar, B.


Marr, D.


Martin, M.

‘Perceptual Content’, in *A Companion to the Philosophy of Mind*.


MacDonald, C.


McDowell, J.
'Criteria, Defeasability and Knowledge', *Proceedings of the British Academy*, 68.


'Singular Thought and Inner Space' in Pettit, MacDowell (eds.), *Subject, Thought and Context*,

McDowell & Petit (eds.)


McGinn, C.


McLaughlin, B.

'Davidson's Response to Epiphenomenalism', in Heil and Mele (ed.), *Mental Causation*.

'Anomalous monism and the Irreducibility of the Mental' in Le Pore, & McLaughlin (ed.) *Actions and Events*.

Meldon, A. I.


Moore, A. (ed.)


Nagel, T.


Nemirow, L.

'Physicalism and Cognitive Role of Acquaintance' in Lycan (ed.)

Owens, J.

'Psychological Externalism and Psychological Explanation',

Papineau, D.


Peacocke, C.


Pessin, Goldberg (eds.)

The Twin Earth Chronicles, M.E. Sharpe, Armonk, New York, 1996

Pollock, J.


Putnam, H.


'The Meaning of "Meaning"', in Philosophical Papers, vol 2, 1975,


Representation and Reality, the MIT Press, Cambridge, Mass, 1988
Quine, W.V.O.


'Quantifiers and Propositional Attitudes', Journal of Philosophy, 1956

Rosenthal, D.

Nature of Mind, Oxford University Press, 1991

Russell, B.


Ryle, G.

The Concept of Mind, Methuen, London 1949, Chapter 6.

Searl, J.


Segal, G.


'Seeing what is not there', Philosophical Review 98, 1989.

Sen, P.K.


Snowdon, P.


Sosa, E.

'Propositional Attitudes De Dicto and De Re', The Journal of Philosophy LXVII, 1970

Stich, S.

Strawson, P.F.


Textor, M.

'The Semantic Challenge to Russell's Principle',

Tye, M.


'The Subjectivity of Experience' in *Mind*, 1986

Van Gullick, R.


Wiggins, D.


Williamson, T.


Woodfield, A. (ed.)


Wright, C. (ed.)


Wright, Smith, MacDonald (eds.)

*Knowing Our Own Minds*, Oxford University Press, Oxford, 1998

Yablo, S.