LOST CAUSES

use and abuse of causality in the philosophy of mind

Michiel Brumsen

submitted for the degree of Ph.D.
October 1999

I, Michiel Brumsen, hereby certify that this thesis, which is approximately 74883 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in October 1992, and as a candidate for the degree of Ph.D. in June 1993; the higher study for which this is a record was carried out in the University of St. Andrews between 1992 and 1997.

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Ph.D. in the University of St. Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Acknowledgements

I would like to acknowledge the help, support and advice without which this thesis would not have been written. It is up to the reader to decide whether it is praise or blame that the people who offered it share with me; however, if it be blame then it should be realised that it is only for the fact that this thesis came into being. The quality of it is my own responsibility, but I am sure it would have been worse without their help.

For first getting me interested in philosophy, I have to thank those who tried to teach me applied physics, somewhat in the manner of the politician thanking the opposition for bringing about his own victory. Jos Uffink, Dennis Dieks, and Jan Hilgevoord played the intentional part in getting me interested in philosophy of physics. Jan Bransen introduced me to philosophy of mind in a most stimulating manner, and years later gave me the opportunity of airing my views during the Human Action and Causality conference held in Utrecht in 1996. Paul Gilbert and Kathleen Lennon gave me a solid base in philosophy of mind, and Kathleen was the one to stimulate me to do a Ph.D.

In St.Andrews, Peter Clark, Bob Hale, Crispin Wright, Leslie Stevenson, Stephen Read, Michael Tye, and John Haldane have all read and commented on parts of my work at different stages. Tim Crane, Bill Child, Stefaan Cuypers, Jaegwon Kim, Brian McLaughlin, Alan Millar, Lynne Rudder-Baker, Paul Snowdon and Charles Travis also argued with and stimulated me. Among my colleagues, Manidipa Sen, Chris Lindsay, and Fiona Macpherson deserve special mention for the feedback they gave me.

Without the financial and moral support of my parents I could never even have started writing this thesis. I would like to thank them for their encouragement and the trust they put in me. The University of St. Andrews Gapper Trust fund also kindly gave me a grant which paid part of my fees for a year. Perhaps even more important than the financial side of things is the encouragement of numerous friends and family, who kept patting me on the back, saying "you can do it!" My sister, holding down a busy job and raising a family as well as writing her Ph.D., is the most vivid proof that it is possible, although I beat her on time.

Roger Squires has been a remarkable supervisor. I thank him for always finding time to fit me in, patiently arguing things through, cheering me up when I felt discouraged, bearing with me when in the true spirit of the student-supervisor relationship I rebelled like an adolescent against his father. Although we did have differences - usually over how tolerant one should be towards philosophers' jargon - I suspect that much of this thesis originated in Roger's thought.

Last but not least, I should thank my colleague Lucie Antoniol, who also became my wife during the work on this thesis. Much of my understanding of what it is to be a philosopher comes from her. A steady homebase, plus all the philosophical input, feedback and stimulation I could wish for 24 hours a day (well, almost), are amongst the things which probably made writing this thesis a lot easier than it is for many others. Lucky me! I hope I can repay her some of this debt of gratitude.

Stirling, 4 October 1997

Note with resubmission:

Many thanks to those who did not lose their belief in me. This includes Peter Kroes, Henk Zandvoort and my other colleagues in Delft who made it possible for me to take time to work on this resubmission; Gregory McCulloch, Crispin Wright and Roger Squires for helpful criticisms and discussions; and Lucie Antoniol for continuing support and encouragement, both of a practical and philosophical nature. And others, whose support was perhaps less tangible, but their encouragement gratefully received.

Delft, October 1999

ABSTRACT: "Lost Causes – use and abuse of causality in the philosophy of mind"

This thesis argues that certain uses of the notion of causality in the philosophy of mind amount to abuse. I start out by making an inventory of the different causal claims that are made. Some philosophers claim that for something to fall under a mental concept such as action or perception, certain conditions to do with causal ancestry or progeny need to be fulfilled. Others claim that we can only understand the strength of psychological explanation by conceiving of it as causal explanation. Yet others claim that the subject matter of psychology consists of essentially causal states. Causal claims are made in a large variety of variations upon the three basic themes named above.

These causal claims are unlikely to be conceptually unrelated to each other, and I try to say something about how they are related. My first concern, however, is with the necessary causal ancestry claim mentioned above. I explain what exactly it amounts to, and consider the arguments for it; then I discuss a number of objections. Once the objections are set up, it is also possible to see to what extent they have impact on the other possible causal claims. I follow this procedure for two mental concepts, namely action and perception.

In this way it also becomes clear that there are important parallels between forceful objections against causal views of action and perception. The most important parallels have to do with scepticism about the 'external world' brought on by a so-called common-factor approach, with deviant causal chains, and with infinite regress. But there are also some disanalogies: the philosophical preoccupations are in the case of action mostly directed on the explanation of action, whereas for perception the main question is how to distinguish it from qualitatively similar states such as hallucination. Consequently, discussions tend to have different emphases. It is my contention that important cross-fertilisation is possible, and that objections against causal views can be generalised to other mental concepts.

Two arguments in favour of the causal ancestry condition on mental concepts are discussed at length in the penultimate chapter. Firstly, the argument from counterfactuals argues that, given the truth of counterfactuals such as "had the cat not been there, I wouldn't have seen it" in cases of genuine perception, there must be a causal link making the counterfactual true. I show that there are good reasons for not taking the counterfactual in question to be a causal counterfactual, in the sense that the argument needs it to be one. Secondly, the argument from natural kinds takes it that mental kinds may well be discovered to be natural

kinds. I argue that even if we could flesh out what such a discovery would amount to, no conclusion about mental *concepts* is available from an argument about the *reference* or subject matter of such concepts.

I conclude with an analysis of what in my view is the basic problem with causalist theories of mind. I locate it in the commitment to representationalism, understood as requiring inner mental states representing the outside world. I show that construals with less ontological commitment are available in order to capture our common-sense talk of people representing things. Dualist philosophy of mind such as Descartes' or Locke's is committed to representationalism as much as is contemporary monist or materialist philosophy of mind. My rejection of causal-ancestry accounts of mental concepts, therefore, is no rejection of monism and (relatively recent) scientific insights, but only of a certain use of the notion of causality in conceptual accounts which philosophers have mistakenly felt compelled to adopt by the advancement of science.

TABLE OF CONTENTS:

1 Introduction

1.1 Causal theories of the mind

The philosophy of mind is in crisis. It has seemed, especially in the last decennia, that on the question as to how body and mind are related we were moving towards a consensus. That consensus consists of a rejection of substance dualism, a doctrine which was perhaps introduced by Plato, and very vividly formulated by Descartes in the 17th century. Descartes held that mind and matter are separate substances and that the mind is in contact with the body through the pineal gland. In the 20th century several different expressions have been given to the idea that there is just one substance, and that the mind is somehow realised in the bodily substance. In the light of various developments in psychology and the brain sciences this has become an attractive position. However, one problem with such materialist positions has proved stubborn: how are we to create, in our metaphysical picture, the space for a causal contact between the mental and the physical? The question that needs addressing is: what kind of causal role, if any, do mental states and events have?

In current philosophy of mind, there is a large variety of claims of a causal character about the mental. The aim of this thesis is to disentangle some of these claims, and to show that the support for some of the most commonly held ones is not as strong as is supposed. It will in fact emerge that these claims, without any further support, do not deserve to be the default position, since the arguments in favour of them do not withstand close scrutiny. And if we cannot take these claims for granted, neither can we, in general, take for granted what causal role the mental supposedly plays, or indeed, that it plays any causal role. The view that will be spun out in the coming chapters is that although some uses of the notion of causality in the philosophy of mind can be quite legitimate, others are not so legitimate and amount to abusing the notion of causality. In this way I hope to make some contribution to the debate over mental causation, which seems to have led the philosophy of mind into crisis.

What type of causal claim forms my main target? Let me begin by outlining what kind of question they are meant to address. I'll take action as an example.

Everyone is familiar with the idea that people perform actions; but does this mean that everybody knows what actions are? Most of us have learned about actions through example, by ostension or lists of action. But even though we all have a pretty good grip on the everyday use of the term 'action', and there are

relatively few problematic cases, the question "what is action?" seems to be a legitimate one. If we develop an understanding of under exactly which conditions the concept ceases to apply, it will be easier to decide on the problematic cases. This is not a matter of the philosopher simply proclaiming authority: the philosopher's job is to painstakingly chart, bring out in the open, and if possible or needed try to make consistent, the conceptual practices already being used by any competent language user. The aim is not to prescribe but to describe and so promote clarity and consistency. Such clarity will be of benefit to questions of psychology, morality, and justice, to name the most important ones.

Whenever a question such as "what is action?" is asked, we need a contrast class: i.e., the implicit second half of the question is, "..as opposed to Y". So what could Y be? Y is generally assumed to stand for 'a mere event'. This brings in the assumption that actions are (natural or physical) events, but also something over and above that. In other words, there is the class of events, and a subclass of it is formed by those events which are also actions. Obviously there are plenty of events that we have no trouble distinguishing from actions, such as volcanic eruptions and solar eclipses, so to make the question more interesting we can substitute for Y a subclass of events which comprises the subclass of actions: that of bodily movements. The question has now become an epistemological one: how can we tell the difference between an action and a mere bodily movement? If one is impressed by this question, one is likely to come to the conclusion that it is at least logically possible that nothing intrinsic to the event can differentiate between action and mere movement. Therefore, a differentiating factor has to be extrinsic to the event, and causal ancestry is a likely candidate: the difference between winking and blinking, for example, is that the same movement is caused in different ways. A similar story can be told about visual perception with mere visual experiences (comprising hallucinations and illusions) as contrast class.

Given the story these theories tell about mental concepts, they could perhaps more precisely be called causal-ancestry-theories. It is important to say what these theories do *not* say, and, consequently, what I am not reacting against. One way in which we use the notion of causation in mental talk is to say such things as: "His pretending to have our best interests at heart made me (caused me to get) angry", or "The stories about political sleaze caused the government's being voted out of office." Although it is unclear how this kind of causation relates to a strict notion of (micro-)physical causation, such locutions seem perfectly legitimate uses of a notion of causation. But they have nothing to do with causal ancestry theories of the mental.

Nor are the claims made by causal ancestry theories about the neurophysiological process taking place when somebody acts or perceives. There is overwhelming empirical evidence, which no-one in their right mind would seek to deny, that when someone, say, stretches out their hand, tiny electrical signals are generated somewhere in the brain, which travel down neural pathways to the relevant muscles, causing them to contract in certain ways. Again, when somebody visually perceives an object, light is reflected by the object onto the retina of the subject, stimulating the receptor cells there present, causing an electrical signal to travel to the visual cortex. All this is - I take it - uncontroversial. Causal ancestry theories, however, hold that unless there is a causal link of the described kind involved, the concepts of action and perception do not apply to the phenomenon in question. Not only is this how things *happen* to work with action: it *could not* be different, because if it were, we would cease to apply the term 'action' to the item under consideration. In arguing that support for such strong claims is lacking, I am not being mystical, anti-scientific or trespassing on the terrain of neurophysiology. I want to deny the conceptual claim, while acknowledging that neurophysiology may well tell a true story.

1.2 The role of empirical science in philosophy of mind

I have already to some extent been dodging the question whether conceptual analysis is altogether immune from progress in empirical science, and I'd better say something about that now. In the traditional picture, metaphysics and conceptual analysis dictate the terms within which empirical science has to work. Conceptual analysis has this privileged position because the truths that it arrives at are analytic. Analytic truths cannot be disproved synthetically; for example, no amount of evidence will make us doubt that 2+3=5, or that the law of excluded middle in logic ($A \vee \neg A$ is always true) is true. If we are to come to the conclusion that we were, after all, mistaken, that should happen on the basis of further analytic truths. In that case, the exercise is to make our system of analytic truths consistent. Therefore, if we come across empirical evidence which seems to be in conflict with an analytic truth, the thing to do is to consider which synthetic judgement can be adjusted, so that the inconsistency disappears.

W.V.O. Quine[1] challenged this picture. He suggested that a sharp distinction between the analytic and the synthetic is a dogma that we have to reject. Rather than thinking of the truth of certain propositions as unassailable, we have to think of our body of knowledge as consisting of a core of virtually certain, well-supported hypotheses, with less stable hypotheses further outward at the

[1] Quine 1953

periphery. When an inconsistency arises, it is easier to adjust a proposition at the periphery, because doing that will require little or no adjustment elsewhere. However, sometimes a piece of evidence is very stubborn, and consistency can only be restored at the cost of many such peripheral adjustments. Alternatively, it may be possible to adjust one more central proposition in order to re-establish consistency, in which case that would be the rational thing to do. Consequently, the truth of the central propositions is not unassailable: if the evidence requires that we adjust them, we should be prepared to do that.

Under the influence of the Quinean rejection of the analytic-synthetic distinction, (mainly American) philosophers' attitude to empirical science changed. If the truths of philosophy are not unassailable, the thought is, then perhaps we can find philosophical truths by doing empirical science? However, even if Quine was right in arguing that the distinction between purely empirical and purely conceptual questions collapses, it merely breaks down the division in two different kinds of questions. But this is no licence to start confusing questions, or to run questions together that are really separate ones. Nor does it provide reasons to think that the same answer will do to two different questions. And this, in a nutshell, may be what has happened in the analysis or clarification of mental concepts: the empirical fact that the neurophysiological basis of action involved certain causal links was taken to settle the conceptual question about action at the same time. But that is a confusion: to acknowledge that the results of our conceptual analysis are not altogether immune to *revision* in the light of empirical facts is not the same as saying that empirical results are straightforwardly the *answers* to conceptual questions. A careful appraisal of whether, and if so: which, conceptual truths about the mental are actually supported by the advances in scientific psychology and the neurosciences is needed, and I hope to go some way towards that goal.

1.3 The variety of causal claims

I have given one example of a causal claim, namely a causal ancestry claim about (the concept of) action. That type of claim is the one I am mostly concerned with in this thesis. But I have also said that there is a large variety of causal claims about the mental, and that I would make an attempt at disentangling some of them. 'Causalism about the mental' is not the name of one, clearly delineated position; I will now try to make the most important distinctions. This will enable me to discard some of the positions that I will not be arguing against. Also, a clear statement of the positions will be of help later on in assessing the reach and

relevance of the various arguments that I shall put forward. The more so, should it turn out to be possible to create a (partial) ordering of positions.
One kind of causal claim about the mental is made by those impressed by the progress that has been made in the empirical brain sciences. They hold that:

(CPI) **C**ausal **p**rocesses are, as a matter of fact, **i**nvolved when mental phenomena take place.

In physics and neurophysiology, we see no problem in talking of causal processes. And it seems obvious from the point of view of current science that physical and neurophysiological causal processes do occur when mental phenomena take place. For example, when I wave my hand certain nerve-impulses travel from my brain to the relevant muscles, causing them to contract. It is perhaps less obvious in the case of mental phenomena such as believing something to be the case, since there is no clear 'outer' occurrence. But it does seem a safe bet that if no processes occurred in our brains, we would not be able to believe things either.
The claim that causal processes of some sort occur when mental phenomena occur is philosophically a weak claim: it states that as a matter of contingent fact, causal processes are going on when mental phenomena occur. It does not say anything about the relation between these causal processes and mental phenomena; and (consequently) nothing about whether such processes must occur in other possible worlds, or in other creatures than us, when mental phenomena occur. That is a matter for debate, on which the current claim does not take any stand. The research into what kind of processes exactly are involved has sparked off a lot of research, and has produced many interesting findings. Which of these findings have to be taken notice of by philosophers of mind is a question to which I shall return. However, that physiological causal processes take place at all I regard as a claim too weak to be of any philosophical interest, and I will not pursue it in this thesis.
A slightly stronger claim is that there is some causal essence to mental states and events; the thought is that, necessarily, the subject matter of mental concepts consists of states that are part of the causal network.

(CE) (Even though the concepts of psychological states may be innocent of causal implications, it may be that) mental states are essentially causally active or acted upon, or 'located in the causal swim'. The states and events we pick out by

means of our mental concepts happen to have a **c**ausal **e**ssence of some kind, even though that is not reflected in those concepts.

It should be seen in analogy with the claim that water is necessarily H_2O, even though it is no part of the concept 'water' that water has the molecular formula H_2O. The underlying thought is that mental kinds are, or turn out to be, natural kinds. I will discuss this claim in chapter 5 where I dwell on the role that the idea of natural kinds can play in causalist argumentation, and I will there conclude that given that this claim has no conceptual implications, I have – like with **CPI** – no quarrel with it. As stated here, **CE** is the denial of the position known as eliminativism, according to which there is nothing to which our mental concepts refer – nothing, that is, of which all the things are true which according to folk-psychology are true of mental states.

Of more interest from my point of view are claims of the type that I gave an example of earlier, about the character of mental or psychological concepts. They are claims concerned with the question as to when something properly instances the concept. These are analytical a priori claims, arrived at by conceptual reflection.
The most general claim of this type is the following:

(VCD) There are certain **v**alid **d**istinctions to be made among mental concepts, which can only be captured by using **c**ausal notions.

For example, it may be that the distinction between *de re* and *de dicto* beliefs is only capturable in terms of causal notions: in order to have a de re thought about an individual I must have been in causal contact with it. Although I have worries about this kind of claim, I will not oppose it in general in this thesis – that would be a continuation of the current project.
More specific claims of this type are concerned with the causal ancestry or progeny of mental states and events such as action and perception. Depending on whether a claim is made about a condition being necessary and sufficient, or only necessary, for application of the concept, and on whether the condition in question is about the causal ancestry or progeny of the state in question, we may distinguish the four following different claims:

(NSCA) A specific **c**ausal **a**ncestry is both a **n**ecessary and a **s**ufficient condition for a state to instance the mental concept X.

This is a very strong claim, which is not defended by many. When the sufficiency claim is dropped we do get a widely defended claim:

(NCA) A specific **c**ausal **a**ncestry is a **n**ecessary condition for a state to instance the mental concept X.

This is the type of claim most central to this thesis. What it says is that if we have a mental state or occurrence which is not caused by the right thing in the right way, then it cannot be an instance of the concept X. The precise nature of the condition obviously needs filling in: what, exactly, is this causal ancestry supposed to be?. It is defended, in one form or another, by many authors; for example John Locke, more recently E.J. Lowe, and perhaps also Donald Davidson about action; and Paul Grice, Alvin Goldman and others about perception.
Jennifer Hornsby defends a claim about causal progeny in the philosophy of action, which in analogy to the previous claims can be stated thus:

(NCP) A specific **c**ausal **p**rogeny is a **n**ecessary condition for a state to instance the mental concept X.

In other words, according to this claim, what is important is not what the state is caused by, but what it is the cause of. (Formulating **NSCP** – which is not, to my knowledge, defended by anyone - I will leave to the reader.)

Then there are claims about the nature of everyday (folk-)psychology and the role of causality and causal explanation therein.

(FPCE) **F**olk-**p**sychological explanation is a **c**ausal **e**xplanation. This is what accounts for the force and success of every-day psychological explanations.

Few philosophers nowadays take issue with this claim; Donald Davidson is probably the most prominent proponent of it. He saw off the neo-Wittgensteinian view – proposed among others by Melden - that action explanations place the action against a backdrop of reasons that makes them intelligible. Part of that view was the idea that since there is a logical connection between reasons and actions, there cannot be a causal connection, and therefore action explanation in terms of reasons cannot be causal. More recent authors who take issue with

FPCE are Arthur Collins[2] and George Wilson[3]. They take the view that reason explanations of actions are teleological explanations, and moreover, that such teleological explanations are not reducible to causal explanations.
Closely related to **FPCE** is the following:

(FPCS) Folk-**p**sychological concepts (such as belief, perception, sensation, etc.) are concepts of **s**tates which have **c**ausal powers, and have a place in a causal network. This is what vindicates the idea of folk-psychological explanation being causal.

Jerry Fodor is a prominent defender of this claim. **FPCS** is mostly conceived of as a thesis about how **FPCE** could be true, because it is felt that some kind of story must be told about that. However, other philosophers (such as William Child, Andy Clark, and Lynne Rudder Baker) stress that **FPCS** demands token-identity of the mental and the physical, and take this to be a reason for denying it, while endorsing **FPCE**. Whether this combination of views is consistent depends, as we shall see, on what one takes to be the relationship between causation and causal explanation.
I shall be contesting both **FPCE** and **FPCS**. They are closely related to the causal ancestry claim – how exactly, will become clearer later on. For now, let me point out that **FPCS** can be read not only as something that has to be true if **FPCE** is to be true, but also as a claim about what makes something into a mental state of a certain kind: namely, the place it occupies in a causal network. This is the position known as functionalism. Notice, that the first half of **FPCS** could be given a metaphysical reading, on which it would say the same as **CE**. However, both the functionalist interpretation – which gives a conceptual reading – and the reading on which **FPCS** vindicates **FPCE** exclude this. The latter does so, because an explanation is a good explanation not only in virtue of what the subject matter of the used concepts happens to be, but in virtue of a certain relationship between the concepts themselves that are used to explain. "The flooding occurred because of the hurricane" can be a good explanation, but not "The event reported on page 3 of *the Times* occurred because of the event on page 5 of *the Tribune*". Given these considerations, I will give **FPCS** a conceptual reading, and reserve the metaphysical reading for **CE**.

[2] Collins 1987
[3] Wilson 1989

The final claim which I shall distinguish here – without the pretence to have achieved a complete classification of causalist claims about the mental – is one about the kind of causation in which mental states are involved.

(NPC) We should **n**ot be **p**luralists about **c**ausation, i.e., mental causation and physical causation are one and the same thing.

This claim will play at the background of all the other claims. Any causalist claim about the mental comes rather cheaply if by causation, in that context, we mean something completely different than the physicist does. As background claim, I will discuss it in section 4.3.

In addition to the different causalist claims distinguished here, I must say something about the form opposition to these claims may take. In strong opposition, one may argue that a claim leads to certain intractable problems, and that therefore it must be false. By contrast, one may take the weaker view that the arguments used to support the causal claim do not stand up to scrutiny. The causal claim in question may, on that line, be true, but it needs further or different argumentation in order to establish it conclusively. Moreover, there may be indications justifying some pessimism about whether such argumentation could be furnished. Although in some cases my arguments will point out definite problems with causalist claims, in general they will be of the latter kind. The most concise formulation of the project in this thesis, then, could be put as follows: to put the ball firmly into the court of those who want to argue that certain causal ancestry theses about mental concepts are true.

1.4 Plan of this thesis

The inspiration behind the organising principle of this thesis is well expressed in the following passage taken from an influential paper on perception by Paul Snowdon:

> "It is believed by some that reflection on many of our psychological notions reveals that they can be instantiated by an object only if some sort of causal condition is fulfilled. Notions to which it has been supposed this applies include those of remembering, knowing, acting for a reason and perception. ... Although the present discussion is primarily of causal theories of perception..., some elements in it may be relevant to the assessment, or understanding, of causal theories for some other psychological notions. The reason this may be so is that the principal argument used to support a causal theory of perception of this sort exhibits a similar form to arguments used to support some causal theories elsewhere and involved in any

> consideration of how strongly causal theories of perception are supported is the task of getting clear about the force of arguments with that structure." (Snowdon 1981, repr. in Dancy, p.192)

In chapters 2 and 3, I start with an exposition of causal claims made about, respectively, action and perception. In the same chapters I also discuss the objections to those claims. Chapter 4 is an attempt to bring out the analogies, but also the disanalogies, between the discussions over causal claims about action and perception, in order to see how they can mutually benefit. Chapter 5 is devoted to tying up loose ends from earlier objections, and considers two of the most central arguments to a generally causalist position, namely an argument from counterfactuals and an argument from natural kinds. The last chapter, finally, locates the root of causal claims in representationalism, which is a view on the nature of the mental shared between modern naturalism or physicalism and dualistic views of the mind, and I discuss the mixed blessings of representationalism. I conclude with a brief assessment of the implications of this thesis for the philosophy of mind, and I sketch the direction in which the discipline should develop.

2 Action

In this chapter, I consider the concept of action as a subject of causal claims about the mental. To start with, I will describe the more prominent causal claims about action found in the literature, and consider the arguments for them. I will classify them in terms of the classification given in the introductory chapter. I go on by stating the main objections to causal theories of action, indicating to what extent each type of claim is affected.

2.1 Some causal claims about action

2.1.1 Volitionism

Most actions involve a movement of one sort or another. For example, in opening the door, switching on the light, greeting someone in the street my hand moves in certain ways. When I say something, my jaw, lips, tongue and chest move in certain ways to emit complicated sounds. The class of actions is not identical with that of the movements, however. Bodily movements can be the result of some outside force (as when I trip and fall), they can be reflexes (as when the doctor taps my knee), they can be tics or twitches, or they can be actions (as when I kick a pebble).

The occurrence of an action typically logically implies that a certain movement or happening occurs, but this implication does not hold the other way around. When we see somebody's body move, therefore, it does not automatically follow that it was an action. It seems that we do not directly see whether something is an action rather than a mere bodily movement. Do we see other people's actions or do we see only the bodily movements that are involved in those actions? When pressed on this point, it seems that we have to admit that we cannot see more than the movements.

Since actions imply movements, but not the other way around, actions must be movements plus 'something else'. The extra component which makes a movement into an action is not something intrinsic in the movement, because the movement itself can be the same whether it is an action or not. It is not publicly observable: if it were, it would be clear which movements are actions and which are not. This leads to the tentative conclusion that actions must have a specific sort of ancestry. An obvious thought is that actions are brought about in a certain way, i.e. they are caused in a certain way. What exactly it is that causes actions, and what kind of causal link it is, is something that then has to be fleshed out.

The type of causal theory arrived at on the basis of this kind of this epistemological consideration is that which formulates conditions on ancestry for the concept of action to be applicable. In chapter 1 I have called this type of claim **NCA**.

The epistemological question can also be asked from the 1st person perspective, that of the agent. It then becomes: "how does one know that one acts?" (or: "how do *I* know that *I* act?") This worry is the counterpart of the previous one. It is not that we see our bodies move and wonder, was this an action of mine or a mere movement of my body? In general, we seem to have no problem in knowing when we act or have acted.

But with a bit of imagination, we can conjure up situations in which this is not so clear. Take a situation where one presses a button to set off some explosives. If the explosion is sufficiently far away, and you have no other way of establishing whether the explosion has in fact occurred, you may wonder whether you succeeded in setting off the explosives: the wire could have broken, or the ignition malfunctioned. However, you can be sure that you did press the button. But we can take the case further: imagine yourself blindfolded, and your hand anaesthetised - and now press the button. If your finger doesn't move - because of somebody restraining it, for example - it still seems that something occurred: it seemed to you that you pressed the button, or less committal, you tried to press the button. Another example is one in which the opportunity to act is denied to the agent: we might wire up your nerves so that, just when it seemed to you that you were pressing the button, your finger moved because of an electrical current applied by me to your motor nerves. Whichever of these (thought)experiments seems to be the more convincing, it corroborates the thought that there are two components to an action, each of which can occur on its own. Only when we have both the internal component (intention, trying, it seeming to the agent that he acts) and the external component (the bodily movement) can we have full-blown intentional action; but then still only if they are related in a certain way. The claim that this relation must be causal for it to be a case of action readily suggests itself. Notice, that when we consider the epistemological question from the 1st person perspective, it is more attractive to regard the internal instead of the external component as the action, when the causal condition is fulfilled. Thus, here we have a claim about causal progeny, which I branded NCP in chapter 1.

We have thus far looked at the motivations for a causal condition for the application of the concept of action. But more needs to be said about what exactly the causal relata are supposed to be. Volitionists offer a way of filling in what the causal item is that brings action-hood into the world, and distinguishes

actions from other events. According to them, the items that cause our bodily movements are called volitions. I will discuss two distinct flavours of volitionism; they differ, firstly, in that one is making an NCA claim while the other makes an NCP claim, and secondly, in what is said about the relationship between volition and action.

According to the 'hybrid' variant of volitionism, an action consists of two ingredients: the volition, and the result. It is in virtue of the relation between the two, namely the volition causing the result, that they *together* make up an action: volition + result = action. E.J.Lowe[4] defends this view, which he holds to be John Locke's view as well. What the volition and the result are supposed to be, is perhaps most easily shown by means of an example. When I raise my arm, my willing to raise my arm is the volition, and my arm's rising is the result. The result should be distinguished from consequences of the action: if, for example, I am standing next to a road and a car stops because I raise my arm, then the car's stopping is a consequence, which was caused by the result of the action (my arm's rising). Notice that the claim made here is, strictly speaking, one about the causal ancestry of a component (namely, the result) of action. Nonetheless, substituting 'action' with 'result of action' still allows us to run the epistemological motivation, and I will consider hybrid volitionism to make an NCA-type claim.

One worry about the hybrid version of volitionism concerns the causal link between volition and result. If the volition is some kind of pure mental act, how can it stand in a causal relation to something physical? This problem of causal interaction between the mental and physical is one of the hardest nuts to crack for causalist philosophers of mind, and will be discussed in section 2.2.5. Another problem, it may be argued, is that volition and action have to be logically related. Can something be 'my willing to raise my arm', unless logically related to 'my raising my arm'? The point returns in 2.2.3.

The other variant of volitionism, which I shall call cause-volitionism, has it that the volition itself is the action, but that it is the action just in virtue of its causing the result of the action. Hornsby[5] is a philosopher who holds this view. To re-use the previous example: when I raise my arm, it is my willing (in Hornsby's terminology, my trying) to raise my arm itself which is the action, provided that it brings about my arm's rising. "Every action is an event of *trying* or attempting to act, and every attempt that is an action precedes and causes a contraction$_i$ of muscles and a movement$_i$ of the body."[6] (the $_i$ indicates intransitive). This

[4] Lowe 1995
[5] Hornsby 1980
[6] Hornsby 1980, p.33

quotation makes clear that we have hear a claim about the causal progeny that an event must have if it is to be an action. Cause-volitionism is an **NCP**-type claim.
A worry for cause-volitionism is the following: how exactly can we distinguish between those effects of a trying that do make it into an action, and those that do not? If I try to raise my arm, but instead my big toe wiggles, or nothing happens besides a registration of my neural activity in the experimental set-up to which I am wired, then why does that trying not count as an action? Presumably there must be some relation between the content or description of the trying and the description of the action's result. How do we ensure that the relation between the contents and the causal relation are congruous? This problem about the causal efficacy of content is returned to in 2.2.5.
The most basic question for both volitionist accounts, however, has to do with the notion of a volition. The burden of conceptually clarifying the concept 'action' lies there; does it enable us to give a non-circular account of what action is, and does it avoid an infinite regress? These points will be discussed in 2.2.1.

2.1.2 Anomalous monism

Another causal claim about action is concerned, not with the condition of application of the concept 'action', but with the nature of explanations of action. When we act, we are usually able to give a certain kind of explanation of our action: we can give reasons why we acted thus-and-so. This is connected with the idea of responsibility for one's actions: if I am responsible, there must normally be a way for me to justify what I did - i.e., to give reasons. 'Normally', because sometimes we act unintentionally. For example, during a game of football in the street I kick the ball, and unfortunately it breaks a window. It would be silly to ask why I broke the window, although it is something I did; for I did so unintentionally - I didn't *mean* to do it. Typically, many different descriptions can be given of the same action. Describing an action in one way, I may perfectly well know that I did it and give an explanation of why I did it, whereas this may not be the case when given another description. It seems therefore possible that what I do is, at the same time[7], intentional and unintentional. I know why I kicked the ball, and can explain why ("I wanted to score a goal"), but I can't give you a reason why I broke the window, other than that I am a lousy footballer. Or suppose I go to the supermarket to buy provisions for a dinner party. Of course I know that what I am doing is buying provisions for a dinner party. Now a friend

[7] Hidden under the surface here is the question: how do we individuate actions? That there are two (or more) descriptions where the action has different properties may also be taken as a reason for saying that, although the space-time coordinates coincide, what we have are two distinct actions. This, as we shall see, is not the way Davidson jumps.

looks at the things I bought and asks me: why did you buy things with a best-before-date of the day after tomorrow only? Well, I don't know - I wasn't aware of doing that. But if she is right it would be silly to deny that I did so. Davidson writes: "an event is an action if and only if it can be described in a way that makes it intentional."[8]

Another type of case in which it seems that we cannot explain or give a reason for our actions is when we act for no specific reason at all. Some things we just do, without giving it much thought; if asked for a reason, we reply "Well, I just thought I would", "I felt like it", "I just wanted to", or something of the sort. Many actions are, in fact, like that, although when we are asked why we performed them we tend to rationalise afterwards. Going to the pub for a pint of Guinness, cooking cauliflower for dinner, standing on your head, stroking the cat, looking out of the window can be examples of such actions. Davidson says about these cases:

> "..it is easy to answer the question, 'Why did you do it?' with, 'For no reason', meaning not that there is no reason but that there is no *further* reason, no reason that cannot be inferred from the fact that the action was done intentionally; no other reason, in other words, besides wanting to do it." (Davidson, "Actions, reasons and causes" (1963), repr. in Davidson 1980 p.6.)

In section 2.2.6 will argue that this response won't do, and that there is a substantial class of actions which cannot be explained in terms of reasons. Bearing in mind, however, that Davidson's reply embodies the standard way of looking at matters I will now show how it motivates a causal theory of action.

When we give an explanation, philosophers want to know how to account for the force of it. If we don't manage to do so, we might question whether it is a good explanation in the first place. Explanations of action in terms of beliefs, desires, motives, intentions etc. are almost universally accepted to be good explanations[9]. Many good explanations are causal: they explain why something happened by telling us what it was caused by. "The car-crash was due to a brake failure", "the bookshelf came down because the screw broke", "the light went out because of a power cut" are such explanations. Arguably, all explanations that tell us why something happened - why some occurrence took place - seem to be of such a form. And isn't this what we want to know about actions as well - why they occurred? "Why did you kick the ball?" "Because I wanted to score a goal". This

[8] Davidson, "Psychology as Philosophy" (1974), repr.in Davidson (1980) p.229

[9] With the exception of eliminativists, who argue that such explanations will eventually - after suitable advances have been made in science - be superseded by better ones, couched in terms of neurophysiology; see Churchland 1984.

seems to be a good explanation in virtue of the fact that it gives us a causal explanation of my action.
The causal claim motivated in this way is about the nature of action explanation: thesis **FPCE** (folk-psychological explanation is causal explanation) in the previous chapter. The steps from here towards the thought that folk-psychological states are causal states (**FPCS**) and the idea that what distinguishes an action from other occurrences is exactly that causal ancestry underlying the causal explanation (**NCA**) are small.

Donald Davidson[10] is the most influential proponent of **FPCE**; as we shall see, there are reasons for thinking that he endorses **FPCS** and **NCA** as well, unlike some of his critics. His view is based on the postulation that if an action was intentional, then the agent will be able to give a justification or rationalisation of his action. Take the following example of someone justifying his action. "I mowed the lawn because I thought the grass was too long", my neighbour tells me. Actually, this is elliptic for a piece of practical reasoning that runs as follows:

(P1) I desire that the grass be shorter
(P2) I think that, unless I mow the lawn, the grass will not be shorter.
(C) I set myself to mow the lawn.

What, now, gives the force of the 'because'? According to Davidson, we should think of the reason-explanation as a causal one. To explain an action is to explain the occurrence of an event: and the occurrence of events is explained by means of causal explanations.
An important argument for Davidson's position is that it enables us to distinguish between true explanations of action and mere rationalisations. One may have a number of reasons for a given action, but only one is the operative reason – the reason for which the action was performed. In other words, the real reason. Other reasons may rationalise or, after a fashion justify, the action, but they have to be distinguished from the real reason. My neighbour may want to be outside because he quarrelled with his wife, want to try out his newly bought lawnmower, want to annoy his neighbours on a quiet Sunday afternoon, or indeed wanted the grass to be shorter. Now, if we want to know which reason explains his action and which, by contrast, merely rationalises it, we need to

[10] The first statement is to be found in "Actions, Reasons and Causes" (1963), together with related papers reprinted in Davidson 1980.

inquire into the truth of certain counterfactuals, such as: "Had he not quarrelled with his wife, he wouldn't have mown the lawn", and "had he not desired the grass to be shorter, he wouldn't have mown the lawn". This shows remarkable similarity with how we treat causal explanations. If we want to know, e.g., whether the explosion is explained by the presence of a gas leak, we ask: had there not been a gas leak, would the explosion still have occurred? An affirmative answer tells us that the gas leak was the operative factor – the gas leak caused, and therefore causally explained, the explosion. By contrast, the operator who happened to have fallen asleep might have been awake and the explosion still have occurred, so his falling asleep does not causally explain the explosion. So, conceiving of reason explanation of action as causal explanation enables us to make sense of the distinction between real reason and mere rationalisation.

Davidson's argument becomes more convincing if contrasted with the position which he was reacting against. That position, held by Melden[11] among others, says that a reason explanation places an action within a context or against a background which makes it intelligible. Somebody's sticking out his hand is explained by citing as a reason that he wanted to signal a turn, because it tells us about the rules and conventions relevant to the action. On this account there is no reference, in the explanation, to a preceding cause. But the problem Davidson found with this account is that it gives us no grounds for favouring one explanation of mowing the lawn over another. They all seem to be equally good, in that they mention an existing[12] context which makes the action intelligible. And this conflicts with our explanatory practice according to which one amongst them is the real reason for which he mowed the lawn.

Another important argument is that, according to Davidson, we don't merely want to understand the action, but we also want to know why it occurred. It's all good and well to understand that somebody's sticking out their hand is signalling a turn, but that does not tell us why he performed the action in the first place. To ask for an explanation of action is to ask for an explanation of why something occurred; explanations of why something occurred are causal explanations, and therefore action explanations are causal explanations.

Davidson's position contrasts with the volitionist approach, in that the latter is centered on a condition on the applicability of the concept of action (**NCA** or **NCP**) whereas the first primarily is a thesis about the explanation of action (**FPCE**). However, he makes a further claim when he says: "an event is an action

[11] Melden 1961

[12] There are resources for a reply here, for the assumption that the different explanations mention equally applicable contexts is dubious. I let the point rest here.

if and only if it can be described in a way that makes it intentional."[13] An action can be given a description under which it is not intentional, but as event particular there must be at least one description under which it is intentional, and caused in the way described. A non-intentional action[14] cannot in the same way be justified by a reason for that action. (This is not to say that there may not be reasons for a non-intentional action: I may, for example, have had good reasons for slamming the door, however I didn't act upon them but as I closed the door the wind caught it, and as a result I closed the door somewhat more emphatically than I intended to. Given this story, the proposition "I slammed the door shut because I wanted to show my anger at you" is not true, since the matching counterfactual "If I hadn't wanted to show my anger, I wouldn't have slammed the door" is obviously not true.)

On the assumptions, then, that describing an action in a way that makes it intentional entails the possibility of it being explained in terms of a reason, and that a reason being causally explanatory of an action entails that the reason causes the action (claim **FPCS**), Davidson also makes claims about distinguishing actions from other events in terms of causal ancestry. Here is a relevant quote:

> We end up, then, with this incomplete and unsatisfactory account of acting with an intention: an action is performed with a certain intention if it is caused in the right way by attitudes and beliefs that rationalise it [note of DD: This is where (*Actions, Reasons and Causes)* left things. At the time I wrote it I believed it would be possible to characterize 'the right way' in non-circular terms]." (Davidson, "Intending" (1978), reprinted in Davidson 1980, p. 87)

So what sets (intentional) actions apart from other, i.e. mere, events, is their causal ancestry. Davidson's slogan in "Actions, Reasons, and Causes"[15], "The primary reason for an action is its cause" (where 'primary reason' stands for the pro-attitude + belief pair), may be read as a commitment to claim **FPCS** (folk-psychological states are causal states). I conclude that there is reason to believe Davidson to be committed to claim **NCA**, as well as **FPCE** and **FPCS**.

Why is Davidson's position called anomalous monism[16]? First, it is a monistic theory because it insists on a monistic ontology: there only are physical events. Event particulars do, however, have different sorts of properties: mental properties and physical properties. When we use mental properties in describing an event, we may call it a mental event. So it is important to distinguish between

---

[13] Davidson 1974, "Psychology as Philosophy", p.229, repr in Davidson 1980

[14] i.e., an action which is not intentional under any description – a contradictio in terminis for Davidson.

[15] Davidson 1980 p.4

[16] The term 'anomalous monism' is from Davidson 1970, "Mental Events", repr. In Davidson 1980.

events and their *descriptions*. Second, it is anomalous because there are no strict or exceptionless laws in which the mental descriptions of events figure. Thus, there are no strict psychological laws; however, this does not exclude law-like generalisations. "When somebody is thirsty, they will have drink", for example, is a lawlike generalisation, but not an exceptionless law: if they happen to be in a desert, or have entered a bet about who can go without a drink for the longest period of time, thirstyness may not result in drinking. There are, according to Davidson, also no strict bridge laws between physical and mental properties or events. However, mental properties supervene on physical properties: "there cannot be two events alike in all physical respects but differing in some mental respect"[17]. Supervenience does not imply bridge-laws, for it remains possible that events with the same mental properties have very different physical properties.
It can be understood that mental-physical bridge laws are excluded by anomalous monism in the following way. Davidson holds that causation is a relation between event particulars, and that for one event to be the cause of another, the events must under some description instantiate an exceptionless causal law[18]. Normally, this description would be the physical description. Given that mental events figure in causal explanations, causal relations must hold between them (although, as we shall see, not everybody agrees on this point). So we have the following picture: there are exceptionless laws between events under a physical description (i.e. couched in the vocabulary of physical properties) but no such laws between the same events under a mental description. With this it would be inconsistent to hold that there are strict mental-physical bridge-laws.

2.1.3 Practical realism

There are philosophers who, although inspired by Davidson, find that he goes too far in committing himself to **NCA** as well as **FPCE**. Child[19] offers such a position. He limits the field of enquiry to those actions that do have reason-explanations. "The aim of the argument was... not to explain how something which is intrinsically non-intentional can acquire an intentional character....that reasons explain actions is our starting-point."[20] Thus, the only causalist claim advanced is one about the character of reason explanation.
An important feature of Child's causalist position is that he disavows the slogan 'the primary reason for an action is its cause'. In giving a reason-explanation we

[17] Davidson 1970, "Mental Events", p.214, repr in Davidson 1980.
[18] See Davidson 1967, "Causal Relations", pp.159-160, repr. in Davidson 1980.
[19] Child 1994
[20] ibid. p122

cite a causal explanatory feature of the subject in order to explain his behaviour; yet we may not be able to identify straightforwardly mental items which do the causing of the behaviour. He would deny the claim that folk-psychological states are causal states (**FPCS**), and therefore the causal ancestry claim is not available either.

Another philosopher who has defended **FPCE** about action is Lynne Rudder Baker[21]. Her book *Explaining Attitudes* is mostly dedicated to showing that holding action explanation to be causal explanation does not commit one to the claim that beliefs (and presumably, other mental states) are (token identical with) brain states. In my terminology, she wants to deny that **FPCS** is implied by **FPCE** – psychological explanations could be causal other than by mental states being physical brainstates that are part of a causal network. Baker, like Child, argues that, for an explanation to be causal, it need not be the case that explanans and explanandum are causally related. She surveys the literature on proposed criteria of causal explanation, and finds them all too strict. However, the worry is, again, that the position she ends up with is not clearly distinctively causal. I discuss this objection in section 2.2.7.

Child is less confident than Davidson that the dispute between the causal view and the neo-Wittgensteinian view (defended by Melden) can be solved on internal grounds (i.e., by merely discussing the question: Are reason-explanations of action causal explanations?) only. The reason why he prefers the causal view is that he thinks, firstly, that it gives better prospects for an integrated theory of all mental phenomena, and secondly, that if we want to be realists about mental phenomena the causal view is the only possibility. I agree with him that it is helpful to discuss the wider picture in making up our minds about the causal theory of action, by considering other mental phenomena. But it is hard to see how the fairly weak thesis that actions are causally explained by reasons (**FPCE**) can be instrumental in formulating an integrated causal theory of mental phenomena. This is because, as I shall argue later, a story would need to be told about what makes the position distinctively causal, once the requirement of token-identity, and explanans and explanandum being causally related, is dropped. Moreover, when we look at what type of causal theory Child defends regarding perception (see p.78), it is not clear how what he says about action and what he says about perception fit into an integrated causal theory. There is a tension here because the causal claims about perception that he wants to get support from are causal ancestry claims. Either his claim about perception must be of type **NCA** – in which case it is not clear how it would form an integrated

[21] for her views on this matter, see Baker 1995.

causal theory with claim **FPCE** about action – or he must defend an **FPCE**-type claim about perception as well. He seems, as we shall see, to opt for the latter, but it does not become clear what exactly such a claim would amount to.

2.2 Objections to causal claims about action

2.2.1 Infinite regress

2.2.1.1 the problem of regress

Actions, according to **NCA**, are those movements that are caused in a certain way. But which way is that? Inanimate things can (be caused to) move without thereby becoming agents ("The water in the river moved fast"). Agents, too - or at any rate, their bodies - sometimes move without acting (if I am pushed over, there is a movement of my body without my doing anything). All movements, we may reasonably say, are caused one way or another.

Many actions are caused by further actions: for example, I start the car by turning the key in the ignition lock, or I make a noise by banging the table. This, then, is one sort of actions: movements (or events) caused by further - more basic - actions. However, not all movements or happenings caused by actions can properly be called actions. The chain must stop somewhere: if my opening the window causes a draught, and that causes the door to slam shut, I did not slam the door by opening the window. Nor is it the case that when I cause you to do something, what you do is somehow my action.

I now want to concentrate on the 'cause-end' of the causal chain. Obviously, it has to start somewhere; somewhere along the line there must be an action which is not caused by yet another action - an action 'simply' or 'primitively' done by the agent. If this were not the case, we would have an infinite regress. Such a regress is problematic for two reasons. Firstly, if for every action that I perform I have to perform another action which is to cause it, and if the same argument can be applied to this other action, then "we would have to perform an infinite number of actions in order to perform any action at all. And, if this is so, we do not act at all, for we cannot perform an infinite number of actions."[22]. Secondly, if we say that all actions differ from mere happenings only by being caused by more basic actions, we are nowhere nearer an analysis of the concept of action. If we cannot eliminate reference to action in distinguishing actions from mere events this just

[22] Moya 1990 p.13

means that we have given an analysis of the concept of action in terms of itself. What is the 'source' of action-hood?

2.2.1.2 Basic actions

What we need is the notion of an action that is at the beginning of the causal chain. These are known as basic actions, introduced in the literature by Danto:

> "(1) B is a *basic action* of x if and only if (i) B is an action, and (ii) whenever x performs B, there is no other action A performed by x such that B is caused by A
> (2) B is a *nonbasic action* of x if there is some action A performed by x, such that B is caused by A."
> (Danto, "What we can do", repr. in Danto 1973 pp.435-6)

We need to consider whether this idea of basic actions is a coherent one; if it is not, we would be looking for, or postulating the existence of, non-existent things. We do have a notion of an action being more basic than another: when I say, 'I did A by doing B' we feel that B is the more basic action. However, it needs arguing that what is thus captured is a causal notion of 'basic', and that it does not lead to a regress.

The 'by'-locution is a tricky one. We ring the bell by pushing the button, start the car by turning the ignition key, fire a gun by pulling the trigger. But also, we dial a number by pressing buttons on the 'phone, insult someone by being rude to them, complete a doctorate by writing a thesis. These are different senses of 'by'. In the first set of examples, we have two different events, one being caused by the other. The bell's ringing is caused by the button's being pushed, the gun's firing is caused by the trigger being pulled, and so on. It is cases like these which Danto had in mind, when giving his definition of basicness in terms of causation. But now consider the second set of examples: here, we don't have two distinct events standing in a causal relation to each other. My pressing buttons on the phone doesn't cause the dialling of a number: it is (constitutes) my dialling a number. Writing a thesis is not the cause of a doctorate being completed: to complete a doctorate just consists in the writing of a thesis. 'Doing A by doing B', in such examples, uses 'by' in what we may call a constitutive or compositional sense[23].

Things get more confusing when we consider locutions like, 'I raised my arm by contracting my muscles' and 'I contracted my muscles by raising my arm'. Or, 'I grimaced by exercising my facial muscles' and 'I exercised my facial muscles by grimacing'. Or 'I spoke by making the air resonate my vocal cords and moving

[23] see e.g. Hornsby 1980 p.68

my tongue and lips in a specific way' and 'I made the air resonate my vocal chords and moved my tongue and lips in a specific way by speaking'. In these cases, both of the two contrasted locutions somehow seem to say something right. But this is puzzling; for which of the two actions on either side of the 'by' is the more basic one? Surely, if the 'by' relation is symmetrical, it cannot be used for a definition of basicness at all. On the other hand, if both of the contrasted locutions seem right due to an ambiguity in 'basic', then which of the two senses should we use to identify the 'source' of agency? Yet another possibility is that the two assertions contrasted each time may both be true, in the same sense, but on different occasions.

The first suggestion - that the 'by' relation may be symmetrical - is unpromising. It is not that when the two 'by' relata are swapped, the assertion still means the same. The point, rather, is that since they don't mean the same, the 'by' relation can't be symmetrical.

The suggestion that the two assertions may both be true in the same sense but only at different occasions is just to say that a context-dependence is involved. We can't tell just from the descriptions of the actions, without knowing more about the situation, which action is the more basic one. However, we shouldn't be confused into thinking that, given two descriptions of an action, we cannot decide which is the more basic one of the two, even though it is in general true that on the basis of a description of an action we cannot decide whether or not it is basic. For example, raising my arm can be either basic or non-basic; in the 'standard' situation it is basic, but if I grab it with my other arm and thus lift it, it is a non-basic action (I raise my arm *by* lifting it with the other). Similarly, lowering one's heart rate is perhaps something a yogi does directly (i.e. as a basic action), whereas most ordinary mortals might be able to do it indirectly, e.g. by listening to relaxing music. But in these cases we cannot invert the 'by'-locution. And it is implausible that some people raise their arms by contracting their muscles, whereas others (or perhaps the same people at different times) contract their muscles by raising their arms.

We are left with the suggestion that we have two different senses of 'by' at work, both asymmetrical, but in these examples in opposite directions. In the causal sense of 'by', it is true that I raise my arm by contracting my muscles: my contracting the muscles causes the arm to rise (i.e., that I raise my arm). As was observed earlier, it seems also right to say that I contracted my muscles by raising my arm. This is so because raising my arm is what I do intentionally: it is what I believe myself to be doing, I can explain why I do it, and I can simply do

it even if I have no notion of the contracting muscles involved in such an action. Following Hornsby, I shall call this the teleological sense of 'by' or 'basicness'.

How do all these different senses of basicness - causal, compositional, and teleological - impinge on what we were after? Let me recap. The assumption of the causal ancestry claim about action was that actions are those movements or events that are caused in a specific way. A vicious regress then loomed: many actions are actions because the events or movements involved in them are caused by further actions. At the beginning of the chain there must be an action which is basic in the sense of not being caused by a further action; at the same time, its causal ancestry must differentiate it from a mere event or movement. It is important, then, to know which actions are the basic ones. Simply tracing the causal chain back will not tell us when we have arrived at the basic action, since within such an enquiry we don't know how to distinguish actions from mere events. Another approach is to try to identify in each case that action which is not performed *by* another action. However, this is not going to give one unique result since there are several senses of 'by', some of which will sometimes give contradictory results. To try to identify that action which is not performed by performing another action, therefore, will not by itself lead the way to the causally most basic action. Prima facie, though, there seems to be no reason why we could not, given this insight, combine the two approaches, and consistently disregard all other uses of 'by' than the causal one in the search for basic actions. This does not smuggle the causal approach into our understanding of basicness: given the project, the causal sense is just the one which is needed.

So now we have identified one consistent notion of 'basic' when talking about basic actions. However, with that conclusion the question whether there are such things as basic actions which we can use for giving an analysis of, or clarifying, the concept 'action' has not yet been answered.

2.2.1.3 Is infinite regress a problem for causal claims?

Gilbert Ryle[24] has posed the following dilemma to volitionists: if volitions are what makes for voluntary actions, how about volitions themselves - are they voluntary or not? If they are, then surely we have a regress, because what would make them voluntary would presumably be further volitions. But if they are not, then how could they make actions voluntary in the first place?

As for the first horn: what makes an action voluntary is, for the hybrid version of volitionism, its consisting of two components: volition and result. If the same analysis is to apply to volitions themselves, interpreted as mental acts, then a

[24] Ryle 1949, chapter 3

volition itself must consist of a volition and a result. The problem is thus not one of following a causal chain backwards (as Ryle appears to suggest), but of having to divide a cause up into cause and effect, ad infinitum. Apart from it being metaphysically implausible that such a thing can be done, we would end up with things like "my willing to will to will raising my arm": it seems unlikely that we can give an interpretation to such (potentially infinitely) nested mental states that fits comfortably into any psychology. However, the second horn of Ryle's dilemma unmasks it as a false dilemma: why should it be supposed that volitions are themselves voluntary? That would assume that the voluntary/involuntary distinction applies to volitions, i.e., that they are actions. But according to the hybrid theory volitions are only *parts* of actions. An action just is voluntary if its result is caused by a volition, and no further assumption needs to be made about the voluntariness of the volition.

According to Hornsby, by contrast, volitions (tryings) are actions – the only actions that there are. However, the problem of regress has no grip on her account. This is because tryings do not cause actions. To object that for a trying to be an action it must itself be caused by a trying would therefore be to miss the point: on Hornsby's account, it is not causal ancestry but progeny which makes something an action. In effect, there is an inversion-trick pulled off here, which does stop the regress. However, Moya[25] makes just this kind of objection. He starts from the fact that Hornsby, in order to avoid dualism, asserts that tryings are physical. Surely, he reasons, if tryings are physical they must involve events which are non-actional, such as neurons' firing. But such events can also take place without there being an action. So after all, the assertion that tryings are physical means that tryings do imply certain happenings - in which case there must be something which differentiates such mere happenings from the tryings which involve them. In other words, we are back at starting point. However, this criticism ignores that what makes a trying an action is <u>not</u> how it is caused, but whether it causes the appropriate movement, i.e., whether it has the right effect. Perhaps it occurs uncaused, perhaps it is caused by another physical event: it does not matter. Whether something is a trying is apparently a brute fact, no matter what physical events it may involve.

A legitimate question especially to volitionists is whether their position clarifies the concept of action at all. Lowe emphatically denies that volitions are actions; but then, what are they? The most obvious line of thought, namely that they are mental acts or acts of will seems to be blocked if they are not actions, and if the

[25] Moya 1990, pp.27-29

voluntary/involuntary distinction does not apply, as was needed to reply to Ryle's challenge. Lowe suggests that volitions are like tryings. However, the sense of 'trying' with which we are acquainted is one in which we (voluntarily) decide to try or not to try to do something. In other words, the common-sense notion of trying seems to be one of an action. If volitions are to have any (non-circular) explanatory value here, and if volitionism is not to fall prey to Ryle's dilemma, more needs to be said about what they are.

Moreover, there is a question for the hybrid volitionist about how basic actions can be the source of actioin-hood. The causalist picture promoted by it is based on the identity of actions with physical events: an extrinsic causal link then makes something an action as well as a physical event. E.g., the light's switching on becomes my action of switching on the light in virtue of being caused by my moving my finger. But on the hybrid volitionist view this is only true for non-basic actions; basic actions have another ingredient besides the physical movement (and are thus not identical to it). If basic actions are so different in intrinsic structure from all other actions, what is it that still makes them actions?

In Hornsby's version, we are in effect confronted with the brute fact that tryings are actions, and so stand at the source of agency. But what does this mean? Hornsby seems to think that it is a mitigating factor that of tryings - contrary to volitions or willings etc. - we already know what they are; we commonly speak about them. But there are at least two good reasons to doubt this. The grammar of trying, in our ordinary usage, does not seem to be the same as in Hornsby's usage; her concept of trying appears to be a technical one. Firstly, in order for someone to be trying to φ, it must make sense to ask, "in what way are they trying - i.e., by doing what?" Often, one tries to do one thing but another results: I try to return the ball to my adversary's court but I hit it in the net instead, or I try to switch on the light in the hall but end up switching it off in the kitchen instead because I mistake the switches. However, in Hornsby's use there is also the possibility that the subject does not do anything at all - no failed attempt is identifiable - but nonetheless he tries. The question, "how does he go about trying to φ?" seems, in this case, to have no answer.

Secondly, we only speak of trying when there is some sort of obstacle or difficulty involved in what we do. Yet, Hornsby's theory requires that every time we raise our arm, turn the page of a book, scratch our ear, we also try to do those things. It is of course not true that failure must be implied if the concept of trying is to apply: I can try to do something, and succeed. But there must at least be some resistance, i.e., some difficulty in bringing about that which is tried. It should be acknowledged that Hornsby is not impressed by this objection. In cases

where there is no difficulty or resistance (she would retort), we may not be aware of trying, and not normally speak of it either. That is to say, although we *normally* imply that there was some difficulty involved when we speak of trying, when the implication does not hold it does thereby not become false that we tried. Surely, if we do something without difficulty that does not mean that it just happened while we did not try to do it at all? However, that is not what an opponent needs to argue here. When there is no obstacle or difficulty, the concept of trying just stops being applicable; it is *neither* true nor false that we tried. What is at stake here is a certain view of the relation between language and what is true and false[26], which I cannot hope to sort out here.

Turning now to anomalous monism, it does seem that the role of cause of action is played by entities that we are already familiar with. To quote Davidson,

> If this account is correct, then acting with an intention does not require that there be any mysterious act of the will or special attitude or episode of willing. For the account needs only desires (or other pro attitudes), beliefs, and the actions themselves. There is indeed the relation between these, causal or otherwise, to be analysed, but that is not an embarrassing entity that has to be added to the world's furniture." Davidson, "Intending" (1978), repr. in Davidson 1980, pp. 87-88

Davidson obviously implies that the beliefs and pro-attitudes are just the things that we are familiar with. No new items are introduced; however, it is postulated that these beliefs and desires that we know so well are in fact token-identical with physical events that enter into causal laws with those physical events that are token-identical with our actions. This makes the account not completely harmless. For it requires that our ordinary beliefs and desires be such things that can be token-identical with physical events. Two possible distortions enter here. Firstly, are beliefs and desires indeed the sort of thing that we can think of as states and events? Secondly, even if that is the case, are such mental states and events physically realised?

2.2.2 Neutral bodily movements

The thesis, central to **NCA** and **NCP**, namely that what makes the difference between a mere bodily movement and an intentional action is its causal ancestry or progeny rests on an important assumption. The assumption is that there is something in common between mere bodily movements and intentional actions: namely, neutral or colourless bodily movements. Or, to put it differently, what I called the epistemological motivation for a causal theory of action is taken at face

[26] At the opposite sides in this discussion we would find H.P. Grice and J.L. Austin.

value. The thought was that on the basis of the bodily movements that we see we cannot decide whether they are intentional actions or mere movements. In practice, of course, we do make judgements of this kind, and mostly the right ones; but the point was that in any particular case we cannot ever be sure, since there is always the possibility that an item of the other category - a mere movement if we thought it was an action, and v.v. - involve exactly the same movement. The point, therefore, is not that in general we do not see the difference between intentional actions and mere bodily movements, but that we do not see it by just looking at the movements. Moreover, that is not because of any limitation on our part, in the sense that it would not help to look better or have better discriminatory capacities - there is simply not more to be seen. In other words, the movement itself is neutral or colourless, because no intrinsic feature of it makes it either an action or just a mere movement.

This is the story as it is commonly told. Does it make a compelling case for the existence of neutral or colourless movements? Neuberg[27] thinks that it does not. His arguments address the question: what is the proper unit of observation? The sceptical reasoning that we can never be sure whether we see an intentional action or mere movement would be blocked if it could be shown that we don't observe neutral bodily movements. What we have is two competing pictures. According to one (the orthodox picture), we observe a neutral bodily movement, and on the basis of its context (what happened just before/afterwards; how people normally behave with this kind of object; whether the movement seemed to adapt itself to the ongoing situation etc.) we infer whether what we saw was a mere movement or intentional action. According to the other picture, we see the movement within its context as mere movement or as intentional action. Similarly, a piece of cheese and the moon may look the same under certain circumstances (i.e. from particular distances and abstracted from the context), but they do not have a type of surface in common on the basis of which we infer, together with contextual information (how far away is this object? does it come out of my fridge?) whether it is a piece of cheese or the moon that we are seeing. Whenever we see a moving agent, it positively requires an effort to see their movements as neutral: we either see those movements as actions of theirs, or as things happening to them. Earlier (section 2.1.1) I said that both an intentional action and a mere movement imply that a (neutral) movement took place; and this is exactly the order in which things strike us. We see somebody act, and on

[27] Neuberg 1993 ,p.30 ff.

the basis of that we can infer and say that their body moved - not the other way around.

One may object to this in the following way. Surely it has to be admitted that sometimes we do make mistakes about whether an agent's movement constitutes an intentional action or a mere movement? And if such mistakes are to be at all possible, must there not be an inference that went wrong, or a neutral observation to which the wrong interpretation got attached? Put differently, if such mistakes are to be possible must there not be some similarity between intentional actions and mere movements, and does that not point to a common epistemic element between the two - a neutral bodily movement? But this does not follow at all. I can mistake a paper bag, moving in the wind, for a cat: and that mistake can be explained by pointing out that the two are of roughly the same size, and move similarly. In giving such an explanation, I merely point out in which respects a cat and a paper bag are similar. To say that I saw something of a certain size, moving in a certain way, which was neutral between a cat and a paper bag, and that I consecutively inferred (for whatever reason) that it must be a paper bag, is to say something different. I thought that I saw a paper bag; as it happens they are in some respects similar to cats; and it turned out that I was mistaken. There is no element in common between paper bags and cats, any more than there is between intentional actions and mere movements. The point will return, and be discussed more in depth in section 3.3.1 about perception.

The arguments in this section do not establish that no such things as neutral bodily movements exist. Such a position would be hard to maintain, for it seems obvious that if we use physics to describe a certain part of what happens when someone acts, that is just what we are talking about. However, it does show that neutral movements are not things that we see and then base an inference upon, and also that neutral bodily movements are not explanatory of the possibility of making mistakes as to whether we see an action or mere movement. It thus undermines the epistemological motivation for a causal theory of action of types **N(S)CP** or **N(S)CA**, and throws doubt upon whether an important element in such theories - the neutral bodily movement - really has the character and function that that theory says it has.

2.2.3 The logical connection argument

An important objection to the causal theory of action, which was at one point in time adhered to by many philosophers, is the logical connection argument. It is an objection to causal theories of action of types **NCA**, **NCP** and **FPCS**. The

thought underlying the objection derives from a Humean principle about causation which is widely accepted. Recall that on the Humean view of causation, there is no necessary link between cause and effect; the only thing we have is a constant conjunction of two types of events, which in us induces the thought that an event of the first type must always be followed by an event of the second type. However, there is no logical connection between cause and effect: nothing in the cause itself implies that the effect must occur. If there is such a logical connection, then what we have is not a case of causation. For example, my being a bachelor doesn't cause me to be an unmarried male: the one *cannot* cause the other because they are logically connected. The charge against certain causal theories of action, then, is that the alleged cause of an action is not logically independent of the action, and therefore cannot be the cause at all. Melden, criticising volitionism, writes:

> "..nothing can be an act of volition that is not logically connected with that which is willed - the act of willing is intelligible only as the act of willing whatever it is that is willed." (Melden 1961, p.53)

It would seem, for example, that there is a logical connection between my having a primary reason to turn on the light and my turning on the light: the primary reason (and the same goes for volitions) is defined in terms of the action.

How does volitionism deal with this objection? "He wills to raise his arm" and "He raises his arm" appear to be logically connected. But, says Lowe, this is not surprising, since the first (the volition) is part of the second (the action), and so what we should expect is exactly that the second implies the first, and therefore that they are logically connected. But it is not a problem: for the effect of the volition is not "his raising of his arm" (the action), but "the rising of his arm" (the result of the action). A similar point would apply to Hornsby's flavour of volitionism. My trying to raise my arm causes the rising of my arm, not my raising my arm; if the causing is successful my trying to raise my arm *is* my raising my arm. However, one may question whether the logical connection has disappeared in this way. For is there no logical connection between my willing to raise my arm and my arm's rising? It is true that my arm may rise without my willing it (somebody else lifts it) or that I will to raise it but it doesn't rise (I am restrained). However, in willing to raise my arm do I not implicitly also will that my arm rises? My willing that my arm rises surely is logically connected to my arm's rising. It may be replied that my arm's rising is not the kind of thing that I can will. But that is an ad-hoc answer in need of further defence.

Donald Davidson denies that the Humean principle must always obtain; this is a consequence of his view of causation as a relation between event particulars,

independent of their description. Given that we can always redescribe events, we may for 'A caused B' substitute 'the cause of B caused B'; but obviously the logical dependency introduced in the latter description does not preclude a causal connection between the two. In Davidson's view, then, Hume's principle of logical independence rests on a confusion between events and their descriptions. Moreover, the relation between 'my wanting to turn on the light' and 'my turning on the light' is grammatical rather than logical. Desires are not like simple dispositions; as Davidson puts it,

> "Now it is clear why primary reasons like desires and wants do not explain actions in the relatively trivial way solubility explains dissolvings. Solubility, we are assuming, is a pure disposition property: it is defined in terms of a single test. But desires cannot be defined in terms of the actions they may rationalise, even though the relation between desire and action is not simply empirical; there are other, equally essential criteria for desires - their expression in feelings and in actions that they do not rationalise, for example." (Davidson, "Actions, reasons and causes" (1963) repr in Davidson 1980 p.15)

Moreover, the event that occurs when I, acting on my desire, eventually turn on the light is, according to Davidson, not strictly speaking the object of my wanting to turn on the light: "If I turned on the light, then I must have done it at a precise moment, in a particular way - every detail is fixed. But it makes no sense to demand that my want be directed to an action performed at any one moment or done in some unique manner."[28] This drives a wedge between the object of a desire and the action that it (together with an appropriate belief) causes: the object of the desire is of a certain type, whereas the action is a token instantiation of that type, with a number of added details fixed which are irrelevant to the type. Unfortunately, this results in another - and particularly tough - problem which I will discuss later, namely the problem of deviant causal chains.

Let me return now for a moment to Davidson's distinction between events and their descriptions, and his view of causation which relies on that distinction. Causation, as I mentioned earlier, according to Davidson is a relation between event particulars. For one event to be the cause of another, the only requirement is that under some description the events must instantiate some exceptionless causal law. Thus the spirit of the Humean principle of logical independence of cause and effect is complied with. As a consequence of Davidson's distinctive account of the causal relation, the logical connection argument, which affects many causal theories of action, seems to have no grip on anomalous monism.

---

[28] Davidson 1980 p.6

The position which I called practical realism, finally, is not vulnerable to the logical connection argument in the first place: for it denied that we can pick out causal items which are primary reasons, causing the action. In other words, the position is not committed to a causal connection between reason and action, and therefore any logical connection that there might be is not a problem.

2.2.4 Deviant causal chains

The problem of deviant causal chains is a tough nut for any causalist defending an ancestry or progeny claim. The problem is that it is possible for an action to be caused by a primary reason that one has, without the action being intentionally performed because of that reason. In such cases of deviant causation we typically want to deny that the agent acted because of the reason; and yet the reason was the cause of the action. I already indicated, in the previous section, the source of the present problem: Davidson, in answering the logical connection argument, drives a wedge between the object of a desire and the performed action. The latter, he argued, is just an instantiation of the more general type which is the object of the desire. If I want to turn on the light, generally that desire does not include details about at which instant I want to do so, or in which precise manner (with my left or right hand, index finger or thumb, etc.) However, this move blurs the distinction between acting and having a reason for it, and acting because of that reason. For it is possible that my having a reason causes my bodily movement in a non-standard way.

Here is Davidson's own, famous, example of a deviant causal chain:

> "A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. This belief and want might so unnerve him as to cause him to loosen his hold, and yet it might be the case that he never *chose* to loosen his hold, nor did he do it intentionally. It will not help, I think, to add that the belief and the want must combine to cause him to want to loosen his hold, for there will remain the *two* questions *how* the belief and the want caused the second want, and *how* wanting to loosen his hold caused him to loosen his hold." Davidson, "Freedom to Act" (1973), repr in Davidson 1980, p.79

What emerges from this sort of example is that the causal condition, which says that the behaviour must be caused by a primary reason, is not sufficient to establish that the behaviour is in fact a case of intentional action. For it to be sufficient, the behaviour must be caused 'in the right way', i.e. via a non-deviant causal chain. The problem is to spell out what this 'right way' is non-circularly,

i.e. without making use of the concept of intentional action in doing so. I will now discuss four attempts to do so.

The first attempt would be to say that the causal chain from reason to action must, in order not to be deviant, constitute the agent's practical reasoning[29]. In deviant cases, the intermediate states in the causal chain from primary reason to behaviour are typically not token-identical with the mental states constituting his practical reasoning. For an agent to fulfil a complex desire, he must first by practical reasoning form intentions to do certain more basic things, i.e., he must make an action plan. When the behaviour fulfils the agent's desire in a way which he hadn't planned, we have a deviant case. Chisholm has a famous example of a nephew who has resolved to kill his uncle. He speeds to where his uncle lives, and accidentally hits a pedestrian, killing him. The pedestrian turns out to be his uncle. In line with the current proposal, we would say that he didn't kill his uncle intentionally: for his plan was not to do so by hitting him with his car - he carried a gun to that purpose. However, whereas the proposal works for this example, it doesn't survive a modified example. Suppose the nephew *did* plan to hit his uncle with his car: does that make his accidentally hitting a pedestrian who turns out to be his uncle into an intentional killing? The current proposal, counter intuitively, answers that question in the affirmative.

Essentially a different formulation of the same idea is to say that the instrumental intention must have a suitably restricted content. This again is the idea of an action plan: an intention with rather general content will be mediated by more direct, basic, intentions. George Wilson argues that any example can always be adjusted to make the proposal fail:

> "It is granted on all sides that the having of an intention i can, in proper circumstances, cause a range of wholly involuntary responses: e.g., blushing, sweating, fainting, and uncontrollable crying. These responses may, of course, include involuntary movements performed by the agent. What is more, it is always possible for such an involuntary movement or series of movements to satisfy, *purely by accident*, the content or a designated part of the content of intention i. (..)Now, no matter how specific, how vivid, the relevant content of i may be, it will still be possible for an involuntary response to satisfy, accidentally, whatever conditions on action are imposed by that content." (Wilson 1989, p.243)

One may still want to object that the causal theorist can stay one step ahead of the counterexamples. We can require that the instrumental intention is concurrent with the action it causes, so that there is, as it were, no room left for deviant causal chains in between. It is not clear how this helps; Davidson's example of

[29] see Bishop 1989, p.134

the nervous climber would pass this test, whereas we feel that it cannot qualify as an intentional action when he lets go of the rope through nervousness at the very intention of letting go. But surely there must be an intention somewhere that is so specific that *only* the non-deviantly caused action fulfils it, excluding all other behaviour? This is the driving thought behind the proposal, but it is mistaken. Firstly, to require that the agent always has such detailed intentions goes too far. Often, one intends to do something without having a very precise idea of how to go about it. Secondly, and this is more important, the idea that the intention must in detail match just the one action is in tension with one of Davidson's replies to the logical connection argument. The charge that the intention and the action are logically related, and that therefore one cannot be causing the other, was replied to by saying that the causing intention is more general in character: a number of different token actions could all satisfy it. An action is always performed *here*, *now*, and *in this specific manner*; the intention doesn't include those details - if it did, the logical connection argument would loom again. To sum up: whereas relaxation of the vividness or immediacy condition opens the floodgates to examples of deviant causation, making it more restrictive runs counter to what was said in reply to the logical connection argument.

The third attempt to exclude deviant causal chains essentially leaves the task to the neurophysiologist[30]. In those cases where the behaviour is full-blown intentional action, it must be caused by the primary reason in some standard characteristic way. It is not, so the argument goes, the task of the philosopher to find out what way that is; what is needed here is not conceptual analysis, but empirical research. The neurophysiologist will be able to find out by what processes the causation of behaviour proceeds in cases of intentional action, and so all dissimilar processes can be classified as deviant. No behaviour caused in the way in which the climber's hand-opening and letting go of the rope, for example, can ever be an intentional action. However, the strategy is risky. How can we justify the conviction that such a typical causal process, characteristic to non-deviant behaviour, does exist and will indeed be discovered? It cannot be justified other than by begging the question at issue, which is: are actions caused by primary reasons? Only if one thinks the answer is 'yes' can one be confident that the neurophysiologist eventually will find out how the causing characteristically takes place. Secondly, on what basis can a neurophysiologist decide whether a specific case is deviant or not? What we are looking for, remember, is a principled way to distinguish deviant from non-deviant cases. The

[30] see Goldman 1967 p.61

most a neurophysiologist can do is to look at those cases which we intuitively classify as intentional action, and try to find common characteristics of the causal processes involved. But this is to work the wrong way around. We (obviously) already know which are the cases of non-deviant intentional action; but we don't know this in virtue of knowledge of the neurophysiological processes involved. There must be some (conceptual) principle at work, which a restriction on causal processes - if we are lucky - can mirror. This undertaking, then, can be compared with trying to find a principled distinction between tables and chairs by looking at their micro-physical constitution. Still, there might be some attraction to this line if it is argued that actions form a natural kind. I will not further pursue this here, but discuss it at length in section 5.2.

The last attempt that I shall consider here is found in Bishop[31], dubbed 'the sensitivity strategy'. Bishop proposes to close the gap between intention and action, in which deviant causal chains flourish, in the following way:

> "In basic intentional action, the agent *carries out* a basic intention by making controlled bodily movements to fit its content. (..) There is exercise of control if and only if the causal link from basic intention to matching behaviour is sensitive, in the sense that *over a sufficiently wide range of differences, had the agent's intention differed in content, the resulting behavior would have differed correspondingly.*" (Bishop 1989, p.150)

Suppose, for example, if the nervous climber had instead formed the intention to hang on to his partner, he might still have let go of the rope due to his nervousness (for he realises that if he does hang on he will go down as well). Then it is clear that it is his nervousness, and not in the straightforward sense his intention, which caused the behaviour - and we must exclude the case because of deviant causation. In the case of the nephew and his uncle, we can ponder what would have happened if the nephew had intended to rescue his uncle from a dangerous situation he believes him to be in. He would still have sped, and hit the pedestrian which turns out to be his uncle - so again the example is exhibited as deviant.

The sensitivity strategy is indeed able to rule out the deviant cases. However, we should wonder what it is that makes this specifically a causal strategy. It seems right that an action, in order to be intentional, must be sensitive to the matching intention, so that a different intention would result in a different action. But a counterfactual dependence of action on intention by no means implies that the intention must *cause* the action. For example, the position that citing an intention in a reason-explanation serves to understand an action by placing it against a

[31] Bishop 1989, p.148 ff.

certain background gives the same result: had the intention been different, the action would have been different. On that position, obviously, the counterfactual is not given a 'cause-effect' reading. I come back to this theme in chapter 5.

Let me try to get clear about what is at stake. The causal thesis is that the difference between mere behaviour and intentional action consists in the action being caused by a primary reason, or an intention. The problem of deviant causal chains is that such a causal link between intention and behaviour is not sufficient to make it into action. The sensitivity condition is an extra condition over and above the requirement that the intention cause the action. To all appearances it works, and captures the point that we were looking for. However, it can do the work alone - i.e., the requirement that there be a causal link drops out as irrelevant. It is sufficient for mere behaviour to be intentional action that there be an intention such that the action is counterfactually dependent on the action. It *may* be that the counterfactual dependence is implemented by a causal link but it does not *have* to be so. In replying to the objection against the causal ancestry claim, the sensitivity condition has made causal ancestry superfluous: it has undercut the motivation for it, which was that of formulating a condition which would draw the distinction between mere movements and actions.

Having considered four attempts to save the causal theory of action by formulating extra conditions to exclude deviant causal chains, and having found none of them satisfactory, it seems that defeat must be conceded. Davidson, who initially was optimistic about finding a solution, has given up hope 11 years after his first seminal paper ("Actions, Reasons, and Causes"):

> Can we somehow give conditions that are not only necessary, but also sufficient, for an action to be intentional, using only such concepts as those of belief, desire and cause? I think not." (Davidson, "Psychology as Philosophy" (1974), repr in Davidson 1980, p.232)

Given such pessimism about the possibility of solving the problem of deviant causal chains, can we just turn our back on the problem? Can we admit that it cannot be solved but hold that this does not really affect the causal theory of action? It depends on what we take the causal ancestry claim to say. If we can be content to have only a necessary (**NCA**), but not a sufficient condition for some event to be an intentional action (**NSCA**), there is no need to worry. However, this does not seem to be in line with the initial question, which was: what distinguishes actions from mere events? The possibility of deviant causal chains makes it the case that the proposed distinguishing factor only does the job if certain other conditions are met - for what we have in deviant cases are not full-blown actions. As we have seen it seems impossible to specify these conditions

in a non-circular way. We should ask ourselves what work the causal condition does at all, when there is a whole range of examples in which the difference between action and non-action doesn't have to do with what the behaviour is caused by. Perhaps the use of the causal condition generates the right answer in many cases just as a matter of luck - which really means that, if any, there might as well be some other principle at work.

2.2.5 The problem of mental causation

2.2.5.1 Dualism and causal chains

Distinguishing between mere bodily movements and (intentional) actions by means of their causal ancestry is all very well, but how do the elements of actions' distinctive causal ancestry (beliefs and other 'pro-attitudes', desires, intentions) fit into the metaphysical picture? More specifically, do such items fit into the metaphysics as having the distinctive character they have, or do they do so in virtue of being token-identical, or even reduced to, other states and events - e.g. physical? The latter question has been much debated throughout the history of the philosophy of mind, and is generally known as the mind-body problem. The question as placed in the context of the causal theory of action has come to be called, in the last decade or so, the 'problem of mental causation'. The worry underlying it is this: can our favourite metaphysical picture of the relation between mental and physical states allow both for the interaction between the two which common sense seems to prescribe, and for having the mental play that role without losing its distinctive character as mental?

Cartesian mind-body substance-dualism is one of the more obvious and well-known positions to run into trouble on this point. If the mind and the body are defined as mutually exclusive substances, how does the mind make the body move? Gassendi put the question as follows: "How can there be effort directed against anything, or motion set up in it, unless there is mutual contact between what moves and what is moved? And how can there be contact without a body...?" We run into the paradoxical conclusion that our bodies must move of their own accord, and thus there cannot be any free will since what we think and will makes no difference to what we do. The thesis that the mental makes no difference, that it is only an idle accompaniment to our actions, goes by the name of epiphenomenalism.

If it were only Descartes' metaphysical picture which led us to worry that the mental may be an epiphenomenon, the topic would not be so much discussed as it is; for Cartesian dualism is taken seriously by few. However, the more

sophisticated answers to the mind-body problem that have been developed since have not made the problem go away. Yablo writes:

> Why should epiphenomenalism concern anyone today? Part of the answer is that dualism is not dead, only evolved. Immaterial minds are gone, it is true, but mental *phenomena* (facts, properties, events) remain. And although the latter are admitted to be physically realized, and physically necessitated, their literal numerical *identity* with their physical bases is roundly denied. (...) Epiphenomenalism has been evolving too; and in its latest and boldest manifestation, this is all the dualism it asks for. (Yablo 1992, p. 246)

The currently dominant metaphysical picture may be called non-reductive physicalism. It comes in different flavours, but whether a version is possible that successfully deals with the problem is a question that I will discuss later on. Jaegwon Kim is one philosopher who has been devoting his efforts to arguing against non-reductive physicalism, on the grounds that there is a problem with mental causation. He points out that non-reductive physicalism is a dualism, albeit a dualism of properties rather than substances, and according to him it runs into similar problems as substance-dualism. Of the several arguments he gives in different places I pick one here.

It goes as follows[32]. The claim that mental properties are not reducible is most plausibly taken as saying that they enter into causal relations of their own. E.g., my being nervous causes me to tremble. What kind of causal relations can mental properties enter into? Presumably they can cause other mental properties; they can cause physical properties ('downward causation'); or they can cause properties one level upward (social?; 'upward causation'). Step one involves seeing that we cannot have same-level causation without downward causation. For let M_1 cause M_2 , and let M_2 to be realised in the underlying physical property P_2. Now there appear to be two answers to the question: why is this instance of M_2 present? We can say: "it is there because it was caused by M_1"; or "it is there because there is an instance of P_2, which realises it". Are these competing explanations? It is not plausible to suppose that M_1 and P_2 are jointly sufficient for M_2 to occur; nor that M_1 and P_2 are causally overdetermining M_2, i.e. that they are pre-emptive. We must then suppose that M_1 causes M_2 by causing P_2, its realisation base: and so we are committed to downward causation. What's wrong with that? Nothing in itself; but we get into trouble when combining it with upward determination. For if M_1 causes P_2, its presence is a sufficient condition for P_2 to occur. But, moreover, since P_1 realises M_1, P_1's presence is a sufficient condition for M_1 - and so by transitivity, P_2. So why not

[32] Kim 1993b

say what we would intuitively have wanted all along, namely that P_1 causes P_2 directly? Now we have made the mental into an epiphenomenon. It is not my being nervous that makes me tremble, it is the neurophysiological state which realises my being nervous that causes pulsating muscle-contractions. The neurophysiological state in question could as well be realising my state of utter calmness, or my thinking that I need a drink - and so, now that it has lost its causal powers, we have no interest in the mental anymore.

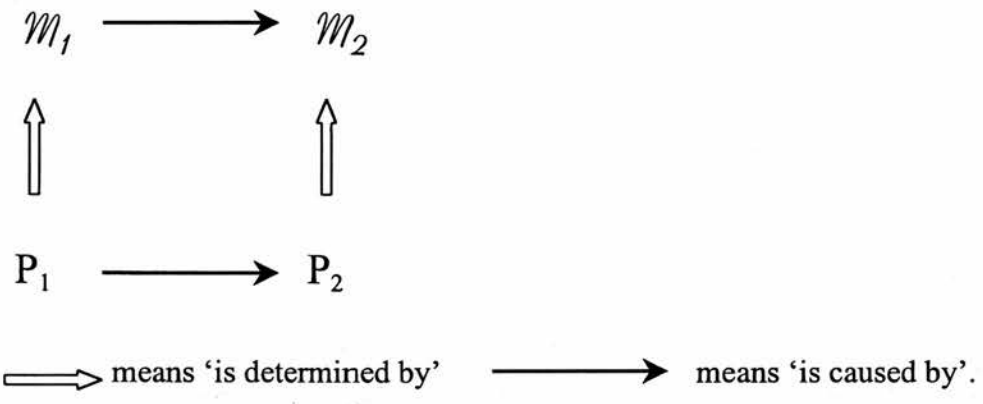

fig. 1: a mental cause and effect, and their underlying physical events.

It is worth noting how Davidson's anomalous monism is affected by this criticism. Although his position is a form of non-reductive physicalism, Davidson would deny that properties on one specific level enter into causal relations, but rather say that event particulars are causal relata just in case some description or other makes them fall under a causal law. This causal law will typically be linking the events described physically. But now his combination of token-identity of the mental with the physical and anomalism of the mental ensures that no laws can be formulated on the mental level, neither can mental descriptions be systematically mirroring physical ones. This, again, makes the mental into an epiphenomenon, since the mental does not cause anything 'qua mental', i.e. the mental description of a state seems irrelevant to what it causes. In Stoutland's terminology, Davidson's anomalous monism gives rise to the mystery of congruence. Somehow, what we desire and intend is (mostly) congruent with the things we do: for example, when I intend to open a door by turning the knob, that normally is just what I do. But how can this be explained, if we hold that this sequence is lawful only under a physical description? Brainstate B causes bodily movement M, according to a causal law; but this is an extensional description, which has nothing whatsoever to do with the meaning of the mental state or intentional action. These are thus irrelevant, whereas it is just the congruence

between the meaningful content of the latter two items which we would like to have explained! Because of Davidson's conception of causation as a relation between event-particulars, it is even *meaningless* to ask in virtue of what property of it one event caused another - the causal relation does not hold between events in virtue of any property. However, it will not do to reject, on the basis of this, the question of whether the content of mental states is relevant to what they cause; for that leaves the congruence between the content of mental state and ensuing action a brute, unexplained mystery.

It has been argued[33] that we now really have two different problems. One is the problem about how mental causes can fit into a world of physical causation. This is not specifically a problem about the mental; rather, it has to do with the completeness of physics, and could be termed (as by Crane) 'the problem of non-physical causation'. The other problem can still arise when a place is found in the causal network for mental events: even if mental events are causes, can they cause what they do in virtue of the content that they have? That is, supposing that mental events can be causes, is what they are about (their content) causally relevant and efficacious? Similarly, one may wonder how the assassination of archduke Ferdinand could have caused the First World War: such political events are after all implemented by underlying physical events, so why does the physical causation not pre-empt the 'political causation'? Even when one is satisfied that the assassination itself was a cause, it may be wondered which properties of it were causally efficacious: is it relevant that the assassination was carried out using a gun, for example, or that it took place at ten past two in the afternoon?

However, although there may be two distinct problems here, we should not think that they are independent. Crane[34] has suggested that we should seek different (dis)solutions to the problem of mental causation and the problem of the efficacy of mental content, on the basis of the following reasoning. If we were to suppose that the mental is type-identical with the physical, the problem of mental causation would be solved. It is simple to see why this is so: if mental states just are the very same thing as physical states, then physical causation cannot preclude mental causation since mental causes are identical with physical causes. But we could still ask the question: did this mental state cause what it did in virtue of its content, or rather in virtue of physical property p of the physical state that it is identical with? To conclude with Crane, however, that this means we can seek separate solutions to the two problems is too quick. Type-identity of the

[33] by e.g. Yablo 1992, and Crane 1998
[34] Crane 1998

mental with the physical can hardly be thought of as a solution to anything, given that there are strong independent reasons (to do with multiple realisation) to think it to be false. And even if we took type-identity to be true, it would be hard to defend it against the charge that it just displaces the problem[35]. Is it not plausible to think of the causal efficacy of content problem as a development of the same problem, in the sense that an acceptable (dis)solution of the mental causation problem might by the same token solve the causal efficacy of content problem? This is consistent with Crane's assertion that the mental causation problem, unlike the causal efficacy problem, has nothing to do with specifically mental properties of mental states, except that they are non-physical states. As Yablo puts it, the type-identity answer to MC "..only relocates the problem from the particulars to their universal features...mental events are effective, maybe, but not by way of their mental properties; any causal role that the latter might have hoped to play is occupied already by their physical rivals."[36] Yablo, then, does not think that the sources of the problems are different, and says so explicitly: "Strangely, philosophers have tended to treat these problems in isolation and to favor different strategies of solution." (p.248) So what we have is really the same problem in different formulations: the mental causation problem is about events, the causal efficacy of content problem about properties. It is rather like the experience one may have in papering a wall: one may smooth out an air bubble under the wall paper, only to find it to have reappeared elsewhere - the problem is to get rid of it altogether.

2.2.5.2 Physicalism: the completeness of physics

According to Tim Crane, what is at play in the problem of mental causation is the inconsistency of the following five theses:

> "(A) Causes have their effects in virtue of (some of) their properties
> (B) There is mental causation
> (C) The completeness of physics is true
> (D) There is no overdetermination
> (E) Mental and physical causation are 'homogeneous'."
> (Crane 1995a, p.229)

He thinks that we should direct our attention at (C), the completeness of physics. What is meant by this is that every physical effect has a purely physical cause.

---

[35] Anomalous monism finds itself in a similar predicament: whereas the monism ensures that the problem of mental causation is in effect solved (causal relata are events simpliciter, which can be given either a physical or mental description), the proposal generates a number of other problems, and still has to answer the causal efficacy of content problem (although that should not be put as 'in virtue of what do mental states cause their effects', since in Davidson's metaphysics this is a nonsensical question, but as 'how come that content of intentions and actions are congruent?).
[36] Yablo 1989, p. 248/9

Physicalism (the thesis that everything in some sense is - i.e. is identical with or constituted by - the physical) was originally motivated by a firm belief in the completeness of physics and the existence of mental causation (understood as the thesis that at least some mental events have physical effects.) For if we deny overdetermination, then mental causes must be physical causes. However, if such reductive physicalism is rejected on the basis that multiple realisation of the mental by the physical is possible, we end up with non-reductive physicalism according to which the 'is' in 'everything is physical' is that of constitution. But now mental causes cannot anymore be accommodated, since they are distinct from - not identical with - physical causes. Physicalism thus gets into trouble with the problem that motivated it in the first place. In other words, the position becomes unmotivated, and we should abandon it.

If we want to abandon physicalism, but not mental causation, we should question the completeness of physics. In particular, how should we define 'the physical'? If we are to avoid such obvious falsehoods as 'the physical is the spatial', we might try to define it as whatever it is that figures in physical theories. But which physical theories? If we take the current ones, the thesis is false: our physics is not complete, struggling to explain certain phenomena, so the entities figuring in physical theories may well have to be added to. On the other hand, if we take those of a future completed physics, then it seems arbitrary to assume that the mental is excluded. Might advanced physics not integrate psychology within it?

Although the point made by Crane carries some conviction, this is only one way of dealing with the inconsistency of theses (A)-(E). I will suggest, in chapter 6, that the belief that there is mental causation is unmotivated; rejecting it need not make us into epiphenomenalists, unable to account for our ways of explaining actions.

2.2.5.3 Determinate / determinable

An important sticking point is the question: are mental and physical causes in competition? According to Yablo, the argument leading to epiphenomenalism has as a premise the following exclusion principle[37]:

(EX) If an event x (or property X) is causally sufficient for an event y, the no event x* (or property X*) distinct from x (X) is causally relevant to y.

[37] Yablo 1989, p.247

There are two ways of denying the truth of this premise. The first would be to say that mental and physical causation are very different kinds of causation, and that they peacefully co-exist. In other words, the claim **NPC** (no pluralism about causation), which is constantly in the background in the debate over mental causation, can be denied. But this pluralist type of answer does not seem satisfactory, however, unless it can be explained why the two sorts of causation 'march in step' (the congruence problem). The second way would be to say that although mental and physical causes are not the same, they are not in competition because they are somehow different layers within the same causal process. I will now focus on a proposal of the second kind, as made by Yablo[38].

Yablo's proposal rests on the determinate/determinable distinction. Consider the following: an object cannot be green and red all over at the same time, but it can be green and coloured all over at the same time. This, perhaps trivial, possibility has to do with the fact that green is a *determinate* of the *determinable* colour. Put differently, being green and being coloured are not competing properties, because green and colour stand in a relation of determinate to determinable. The importance of this relation to the problem of mental causation is that a determinate and a determinable are never competing causes. For example, suppose that extreme weather caused a delay in a postal delivery. If we are then told that in fact it was a snowstorm which caused the delay, do we thereby have competing causes? That would seem an absurd position to take; for if that is so, these causes must also compete with there being a snowstorm with 5 inches snowfall, or there being a snowstorm with wind force 9 and 5 inches snowfall as potential causes - and so on. The reason why such properties as being extreme weather and being a snowstorm (etc.) do not compete, is that they are related as determinable to determinate. So if mental and physical properties are related to each other as determinable to determinate, they will not be competing for causal relevance, and our puzzle will be solved.

The 'if' in the previous sentence is an important one. But before dwelling on it, I will consider two other potential problems to the Yablo's proposal. Firstly, it is easy to find examples of a determinable being quite insufficient as a cause, with causal relevancy only coming in with the determinate. That some meteorological phenomenon took place is not by itself sufficient to cause the delay in the delivery; it may have been just a band of clouds passing, or a 5-minute spring shower. Similarly, it is the fact that a rock dropped on my foot which caused a bone to break - not the fact that an object dropped on it, for then it may as well have been a twig or a leaf. It is, of course, still true that had the meteorological

[38] Yablo 1989

phenomenon not taken place, the delivery would not have been delayed, and had an object not dropped on my foot, the bone would not have broken. One may doubt, however, whether this amounts to causal relevance, in the sense in which we want contents of mental states to be relevant. Secondly, is it enough to be reassured that mental and physical causes are not in competition? Perhaps physical causes do not pre-empt mental causes - but is what we want to accommodate not the idea that the important, relevant causes are the mental ones? Peaceful co-existence of mental and physical causes goes only half the way. Yablo anticipates this objection, and deals with it as follows. Any mental state can be multiply realised. Therefore, any mental state's physical realisation could have been a different one, and so the particular one that did occur was not necessary for the effect to occur. Mental states / properties are thus more proportional to the effects caused, and they are what we would normally think of as the relevant cause.

Most critics of Yablo seem to think that, on his account, the mental does not cause in the very same sense as the physical does. Since he does not deny that there is a complete physical story, how could the mental and the physical cause in the same sense? Yablo is indeed careful throughout to distinguish causal *sufficiency* from causal *relevance* - the determinates are causally sufficient, whereas the determinables are causally relevant. Both can be said to cause, but since in the case of mental causation mental properties are more proportional, those are the ones we generally pick out.

It seems to me that the criticism that mental and physical causation are 'inhomogeneous' is unfair. Just as determinate and determinable are relative terms, sufficiency and relevance are relative in Yablo's story. 'Being a belief that the roof leaks' is a determinable of determinate physical property p, but it is itself a determinate of 'being a belief'; and we may say that my belief that the roof leaks was causally sufficient, and my belief causally relevant. Given this relativity, the charge that on Yablo's story mental causation is fundamentally of a different status than physical causation is misconceived.

However, is it true that the mental is related to the physical as determinable to determinate? Yablo's thought is that supervenience of the mental on the physical is just the sort of one-way determination of the determinable-determinate relation: "something has a determinable property iff it has some determinate falling thereunder." (p.256). He grants that on the traditional view there must be an asymmetrical *conceptual* entailment from determinate to determinable as well; but "since it is only the metaphysics that matters to causation, we should discount the traditional doctrine's conceptual component and reconceive determination in

wholly metaphysical terms." (p.253). It seems, however, that Yablo is begging the question in the following way. What reason do we have for saying that the exclusion principle does not apply when two properties are related as determinable to determinate? To start with, if we were to endorse the exclusion principle, "almost whenever a property Q is *prima facie* relevant to an effect, a causally sufficient determination Q' of Q can be found to expose it as irrelevant after all." (p.258) But the examples he uses in order to show this, are those in which the conceptual entailment does hold: e.g., is a spot's redness causally irrelevant if its being scarlet is causally sufficient for being pecked at by a pigeon?

According to Yablo, we should conclude that the exclusion principle **EX** is badly overdrawn, and that it should be revised. "Even without hearing the details, we *know* that the corrected principle does not apply to determinates and their determinables - for we know that they are not causal rivals." How do we 'know' this? When we substitute 'physical properties and the mental properties which they realise', we 'know' as well that the principle does not apply - but we do not know why that is so, and this is just what we seek to justify. Unless there is some independent justification, it is a rather *ad hoc* piece of 'knowledge'. Yablo continues: "This kind of position is of course familiar from other contexts. Take for example the claim that a space completely filled by one object can contain no other. Then are even the object's *parts* crowded out? No. In this competition wholes and parts are not on opposing teams; hence any principle that puts them there needs rethinking." (p.259) However, it is a *conceptual* truth that the parts of a material object are where the whole is - and that, therefore, in this case determinable and determinate are not competing. The fact that they are not competing might, for all we know, derive solely from the conceptual component of the determination relation, in which case the metaphysical component is irrelevant.

So we cannot assume that the fact that there is no conceptual entailment between mental and physical properties is irrelevant to the question of whether the exclusion principle applies. Yablo argues that it is irrelevant as follows: "...there is no conceptual entailment .. from the tea's micromechanical condition to its high temperature, yet this occasions little skepticism about the role of the tea's temperature in its burning my tongue." (p.260) The analogy with temperature causation is a good one: just as we are not skeptical about the role of the tea's temperature, we are not skeptical about the role my beliefs play in what I do. But what needs to be done is to show how our metaphysics allows for this confidence. The analogy, therefore, shows that we have the same problem in

another domain ('the problem of temperature causation'), and cannot be used as an argument that the lack of conceptual entailment is irrelevant. The metaphysical may be all that matters to causation; but the problem of mental causation is the conceptual problem of showing that mental causation presents no problem given how we conceive of mental properties, physical properties, and how they are related. Yablo has not, I conclude, solved that problem.

2.2.5.4 Teleological underpinning of content

One way of reformulating the problem of congruence is this. Given that any mental event can be realised by a physical event in a number of ways, how can it be explained that any one out of a group of physical states causes a physical state in another group, where the only thing which the physical states within such groups have in common is that they are different possible realisations of the same mental state? Within physics we find no basis on which to group these events together - if one were to formulate a physical property that they share, it would be a highly disjunctive, or *shapeless*, property. Should we not be able to tell a physical story about why the causal relations between the various physical realisations of mental states run in parallel? In other words, should there not be a story about why such apparently disparate physical cause-effect relations mirror one simple mental cause-effect relation? The naturalist teleologist (Millikan 1984 and Dretske 1988 are the most prominent examples) proposes to deal with this puzzle as follows: with the help of evolutionary biology physical events can be grouped together as having in common a natural function. What seemed to be shapeless physical properties are given shape because of common natural functions: e.g., these physical events all realise hunger, since they tend to produce an event of the organism feeding itself. In this way, the fact that the mental and the physical 'march in step' is explained. In effect, then, the mental seems to be reduced to the biological; this is causal-role functionalism with a naturalistic flavour, where the causal roles get their respectability from realising natural functions.

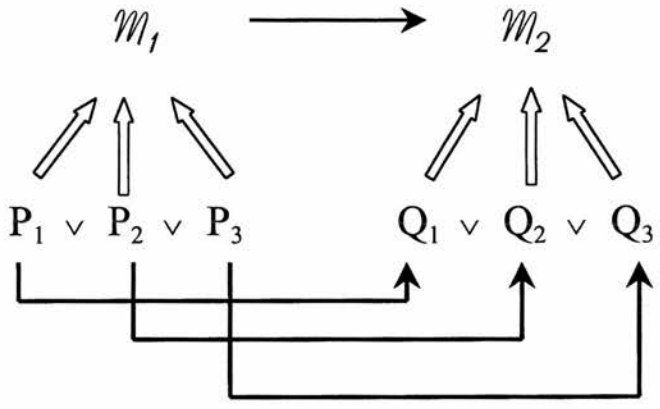

fig. 2: Multiple realisation

This account sounds too good to be true, and indeed it is. To start with, the strategy is a doubtful one, for it has to face the following dilemma. Is biology reducible to physics or not? If it is, then - since reduction is transitive - the mental is reducible to the physical, whereas we agreed that it wasn't. Biology must therefore - this is the other prong - be irreducible. However, if this is true, 'biological causation', like mental causation, is pre-empted by physical causation: the biological is epiphenomenal, just like the mental. Either way, we can't win. At the root of the problem is the fact that we want to regard the physical as primitive in terms of causation, because of its causal closure. For this means that only a *physical* explanation of the congruity of the causal properties of physical events with the mental events that they happen to be token-identical with will do. This demand is obviously not met if we take recourse to another science than physics, like biology.

A more specific problem is connected with the evolutionary conception of teleology. A token state or event only has a function in so far as it is an instantiation of a type of event. The history of a type of event is, on the evolutionary conception, the determining factor regarding function. Previous tokens of this type have brought about g; g is advantageous to the organism; therefore natural selection favours organisms which have, or tend to produce, this type of mental state, which has the function of bringing about g. Obviously we can only tell our story about natural selection with respect to types; speaking of tokens being naturally selected makes no more sense than speaking of the time of day on the sun. A type gets naturally selected over several generations, because it confers advantages of survival to the individuals that have it and is therefore more likely to be handed down to the next generation. So how do we decide what

type tokens fall under? The most obvious way would be to look at the function of the token: but, as I tried to explain, I don't see how we can attribute functions to tokens given the evolutionary conception. This is a serious problem, since the project was exactly to explain why physical state tokens are grouped under mental property types.

Another form of the same problem, which has been stated by several writers, is that the first token contentful state, behaviour or whatever, cannot have a function on the evolutionary theory. Yet it serves the first creature that has or instantiates it equally well as its offspring; indeed, that's why eventually the feature gets selected by natural selection. Or suppose we create a doppelganger with the same brainstates etc. as its original: since it isn't produced by natural means its mental states etc. don't have evolutionary function, therefore are contentless. This is as counter-intuitive as holding that an atom-for-atom duplicate of my bicycle isn't a bicycle, because it hasn't been made in a bicycle-factory.

Millikan[39] does offer another way of ascribing function to token states. Beliefs (to restrict our discussion to these particular items for the moment) are purposeful in as far as they are generated by purposeful belief-forming-mechanisms (BFMs). This would seem to tackle two problems at once, both of which have to do with the fine-grainedness of content. Firstly, some beliefs occur only once, but still they would appear to have contents; secondly, the purposefulness of some beliefs (e.g., the belief that a liking for pea-soup reveals a weakness of character) on their own is hard to spell out. However, nothing gets solved in this way, the problem gets only pushed back: for how can these token BFMs have evolutionary functions?

Purposefulness of behaviour transmits, on the view discussed here, to purposefulness of the mental states that produced it, which transmits to purposefulness of the mechanisms which formed those mental states. If we have trouble attributing purposefulness to any item, we just look at what it produces, or what produced it, and presto! But that can't be right. A thermostat can be the result of some cosmic accident, and still regulate the temperature perfectly. On the other hand, a purposeful mechanism can quite obviously produce something without any purpose: it can be its purpose to make purposeless things.

Dretske [40]has a more subtle approach to the problem. A state, on his view, represents or means that F if it has been assigned the function to indicate that F. A type of internal state can acquire a function by being naturally selected.

[39] Millikan 1984

[40] Dretske 1988

However, this doesn't make the meaning of that state causally efficacious: the function - and therefore the meaning - of the state is assigned to it by natural selection, and not the organism itself. To understand how meaning can make a difference, we have to look, says Dretske, at how organisms learn things. If an organism is sensorily suitably endowed, it will have states that are reliably correlated with certain features of its environment. For example, internal state R will indicate that there is an F in the organism's environment. As it happens, a good way of reacting to Fs is to exhibit behaviour B, and such behaviour provides the organism with positive feedback (e.g., F is the organism's typical food, B is eating it, and satisfying the hunger is positive feedback.) Because of this positive feedback R 'gets recruited' to cause B; when this happens, it becomes R's function to indicate that F, in other words (by Dretske's definition) R comes to represent or mean that there is an F. The fact that R reliably indicates that there is an F is the reason for which it gets recruited to cause B, and therefore R's content is causally explanatory of B being performed by the organism. On Dretske's story, then, two things happen *at the same time* in learning: a state comes to mean something *and* it comes to cause something else. R means that there is an F because it is its function to indicate F; it is its function because it is recruited to cause B. The claim is that the causal explanation that we wanted has now emerged: R causes B in virtue of its meaning that there is an F.

There is a number of worries about Dretske's position, such as: are these causal explanations not circular?[41] Are they applicable to the whole range of behaviour that humans exhibit, or only behaviour that is learned and/or rewarded with positive feedback? Discussing these would take more space than I want to devote to it here, and I will therefore concentrate on the following question: is Dretske's account sufficiently different from Millikan's not to be vulnerable to the same objections? Firstly, mental meaning is obviously reduced to biological facts, and thus Dretske has to confront the dilemma about reduction that I put to Millikan as well. Secondly, there is again the question of how *token* mental states can have meaning, since Dretske's account again is based on evolution. It is *types* of mental states that get recruited to cause behaviour; how do we decide whether a token state falls under a certain type? Not by looking at what behaviour it causes; for that's what is supposed to be explained by the state's content, i.e. its type. My type/token objection just reflects that, once it is agreed that the mental is not reducible (and this is what generates the problem of causal efficacy) no reductive account of mental states, and more specifically of the causal efficacy of mental content, can be made to work.

[41] see Baker 1995, pp.57-62

2.2.5.5 Conclusion: mental causation

Let me try to recapitulate this section on mental causation. The problem arises when we reflect on the place of mental causes in a world of physical causation, as we should if we want to assert that what differentiates actions from mere movements is that they are caused by mental events. The problem of the causal inefficacy of content is a further (not independent) problem, and the problem of congruence is a different variety thereof. This way of organising matters is significant, because it exposes many clever 'solutions' to the problem of mental causation (such as anomalous monism) as being mere reformulations designed to shake of whatever grip the problem has, only to succumb to a different variety or further version of the same problem. I examined two influential proposals to deal with this kind of problem. The first was Yablo's determinate/determinable approach, which although it offered seductive analogies failed because mental and physical properties cannot be conceived of as being related in that way - at least not without begging the question. I also looked at attempts to underpin the causal efficacy of content teleologically, and concluded that these simply do not deliver the goods. The proposals found in the literature are here by no means exhausted: some have called for a reconsideration of the multiple-realisation objection to type-identity theories (Kim), others propose various forms of causal pluralism to ensure that mental and physical causes are not, after all, in competition. This is not a thesis about the mental causation problem, and thus I will leave the topic with this non-exhaustive survey. The point that is important to my overall argument is this: the problem of mental causation is a serious and hard problem, on which philosophers are far from reaching agreement. The claim that there must be mental causation is essential to causal ancestry claims. So it is a problem that my opponents have to face.

2.2.6 Arational actions

The claim that reason explanations of action are causal explanations (**FPCE**) does not by itself require that all actions are done for reasons. However, as we have seen Davidson combines it with the idea that an event's being susceptible to a reason explanation is what makes it into an action: "an event is an action if and only if it can be described in a way that makes it intentional."[42] Thus the claim about the nature of action explanation transforms into one about a necessary

[42] Davidson, "Psychology as Philosophy" (1974), repr.in Davidson (1980) p.229

condition for application of the concept. But are all actions indeed explainable in terms of reasons?

Hursthouse[43] makes a case that there is a substantial class of intentional actions which are not performed for reasons. Many actions are performed with the agent only having an appropriate desire, but no matching belief - and therefore no complete primary reason. Put differently, the agent's action is not the result of a piece of practical reasoning, because one of the essential ingredients is missing. Examples of such actions are e.g.: rumpling the hair of the person one loves, tearing up the picture of the person one hates, throwing an un-cooperative can opener on the floor, shouting at a car that refuses to start, hiding under the bed sheets from fear: in short, 'actions performed in the grip of an emotion'. Hursthouse, for obvious reasons, proposes to call these actions a-rational. Not that the agent has no desires and beliefs at all that are somehow relevant to the action: but he doesn't have a desire and a belief that fit the pattern of practical reasoning which leads to an action. For example, when I shout at my car, I do so because I desire to start it, but obviously I do not believe that I will get it started by shouting at it. Or, I may desire to be safe from danger; but I do not believe that huddling under the bed sheets will be of any help in fulfilling that desire. A-rational actions are pointless, i.e. they are not done in order to achieve anything. In doing them, an agent has no (further) goal and this is why no practical reasoning is applicable. Two possible rejoinders are anticipated by Hursthouse, both of which try to identify a suitable belief-desire pair.

Firstly, someone who wants to save the belief-desire model may hold that in the examples given the agent acts *in order to* express his emotion. Ascribing such a goal to the action generates a possible explanation: "the agent desires to express her emotion, and believes that whatever she is doing *is* expressing it."[44] This move is unsuccessful because such an explanation, although perhaps true in certain cases, is often just false. It is false because the agent is not open to correction: if you were to point out that shouting at my car is not an appropriate expression of frustration, that would not stop me doing so. It is as if you were pointing out that some belief of mine were false, where I don't in fact hold such a belief at all. The argument is not that one may not perform the same type of action *with* such a belief in mind - my psychotherapist may have convinced me that I should express my emotions more rather than bottle them up, and I thus actively seek to do so - in which case the belief-desire explanation is true. But this is not always - probably not even typically - the case.

---

[43] Hursthouse 1991

[44] Hursthouse 1991,p.60

Secondly, one may be tempted to explain 'a-rational' actions by saying "he did that because he wanted to." Of course this is only to cite a desire - but we often leave the relevant belief out in our explanations of action when it seems relatively trivial. For example, we can say "I went to the supermarket because I wanted to buy sugar", and not mention my belief that I can buy sugar in the supermarket - this being a widely shared standard belief. Similarly, when we say "he did that because he wanted to" we might add "and he believed that doing so would fulfil his desire to do so". In other words, the goal of desire-satisfaction would be ascribed to the action; the agent thus performed the action with a reason, but not with any *further* reason. But such a move is misguided. The cited 'belief' is simply a truth about the nature of desires; everybody who understands what desires are knows this, and thus the 'belief' is redundant to the explanation, adding nothing at all to what has already been said. This is different from the belief being implicit because trivial, as in the example of going to the supermarket to buy sugar. Nor can the account be saved by citing the general belief that satisfying the desire would yield pleasure. For most cases this does not work: I do not believe that giving in to the desire to throw the can-opener out of the window, shout at my car, or huddle under the sheets will give me pleasure. In many cases it is the contrary, such as in Davidson's own example: "a man may all his life have a yen, say, to drink a can of paint, without ever, even at the moment he yields, believing it would be worth doing."[45] Perhaps, as Hursthouse remarks, "The only pleasure the agent believes in is "the 'pleasure' of desire-satisfaction", and this is an entirely formal and empty concept of pleasure." (p.63)

Can we not bite the bullet and admit that, given that there are actions not done for reasons, some actions are non-intentional? (not *un*intentional, for those are the actions that, under some other description, *are* intentional.) This would amount to admitting of a category besides intentional and unintentional actions, which is the line taken by Chan[46]. He thinks that actions performed in the grip of an emotion are a subset of non-intentional actions; other cases are mannerisms (tugging one's ear-lobe), and actions performed out of routine or habit. The difference between the two accounts is simply due to a disagreement over what makes an action intentional. For Hursthouse, an action is intentional whenever one is aware of performing it, and it is performed voluntarily; Chan on the other hand seems to regard these as criterion for something's being an action, and for him the *further* question of whether the action is intentional depends on whether it is performed for a reason. In reply to Davidson, Chan denies that an action

[45] Davidson 1963, p.4
[46] Chan 1995

must be intentional under one description or another, whereas Hursthouse denies that an action's being intentional consists in there being an explanation in terms of reasons. Both positions make trouble for the idea that an event is an action iff it is explainable in terms of a reason.

The Davidsonian position, in effect, assumes that three classifications coincide: action/non-action (mere bodily movement), intentional/non-intentional, rational/arational. But there are good reasons to think that there are non-intentional actions; arational intentional actions; and so on. Another way of looking at the matter is perhaps to say that, given that there are several distinctions of interest to be made here, the question of whether a given movement is an action or a non-action doesn't admit of a determinate answer, since it does not specify whether our prime interest is in intentionality, rationality, or perhaps something else. We can of course do some conceptual legislation here, but then the answer will not be particularly interesting. Rather, there seems to be a range, from full-blooded rational intentional action (e.g. purchasing a train ticket) through a-rational intentional action (throwing an unwilling can opener on the floor), a-rational non-intentional action (tugging at one's earlobe), down to mere movements. There may be other distinctions to be made (e.g. the distinction between indirect and basic actions that was discussed earlier; between bodily movements originating from inside or outside the agent; between an action acquiring a special meaning in virtue of social norms and rules; and so on), and maybe there should be a place in the range for unintentional actions.

2.2.7 Causation and causal explanation

The position which I called practical realism embraced the idea that action-explanation in terms of reasons is causal, while rejecting that it is so because of a causal relation between explanans and explanandum. However, how are we to flesh out the notion of A causally explaining B if not by a causal relation holding between A and B? William Child points out[47] that, for one thing, there are many causal explanations made true by underlying causal processes, where the explanans and explanandum do not neatly pick out events that are causally related. Examples are, "the shopping trolley moved because I pushed it", "the ice cream melted because it was in the sun" - what we have here are ongoing processes. The argument here is that "I went to church because I wanted to please

[47] Child 1994 p.108

my mother" might be like that: i.e., it might be a causal explanation without implying that 'I wanted to please my mother' picks out an event neatly.

Another way of looking at the matter, which I find harder to get to grips with, comes from Strawson. According to Child, he holds that "causal explanations are united not by their dependence on a natural relation of causality, but rather by the fact that they are all explanations of the occurrence or persistence of particular events or circumstances, or of general types of event or circumstance"[48]. This appears to be a claim to the effect that what makes an explanation causal is its form. However, if we see the thesis that reason explanations of action are causal in this light, it becomes hard to see why it would be a distinctively *causal* position. It rather seems that from the thesis that reason explanations of action explain why something occurred, it follows *by definition* that such explanations are causal. It makes the class of causal explanations, it seems to me, unacceptably broad. Teleological explanations, too, like causal explanations, can explain why something happened or occurred. When we say that the heating switched on in order to maintain a certain temperature, we explain an occurrence teleologically. It surely is a proper explanation: nobody acquainted with domestic heating systems would argue that it is an accident that a certain temperature is maintained.

It is, of course, the case that the heating's switching on was caused by the thermostat, and we can give a causal explanation to that effect: "the heating switched on because the bi-metal strip in the thermostat made the mercury switch tilt and conduct electricity, which in turn opened the gas supply of the boiler". The causal link shows us in this case how the teleological organisation is implemented; but that does not mean that the teleological explanation reduces to a causal one. One reason for thinking this not to be the case is that causal and teleological explanation are independent[49]. That is to say that we may know the causal explanation of some event's occurrence and be completely in the dark about the teleological explanation of it (if there is any). Conversely, we may know the teleological explanation of some event without being able to causally explain it.

However, can reducibility not be true without our knowing it? In the case of metaphysical reducibility this is true: chemical properties and states, e.g., were reducible to physical properties and states before we knew this to be the case. But explanation is an epistemic matter: something is an explanation relative to our beliefs and knowledge. Therefore if we do not know one kind of explanation to

[48] Child 1994 p.100
[49] see Collins 1984 p. 351

reduce to another, then it doesn't. Once we knew more about the reduction of chemistry to physics, we could explain certain occurrences that we formerly explained chemically by giving an explanation in the terms of physics. For example, the chemical notion of valency of an element, which was used to explain why certain elements react with each other and why they form the numeric combinations that they do, was shown to reduce to the physical notion of number of open spaces for electrons in the outer shell of the atom. However, arguably the result is two different (but both of them true) explanations of the same thing, not a reduction of one explanation to the other.

The question as to whether teleological explanations reduce to causal ones is a difficult one which I do not intend to answer definitively here. My aim was merely to make plausible that the causalist's premise that all explanations of occurrences are causal explanations is in need of more defence. But my opposition to the mere **FPCE** claim is more substantial than that. In section 5.1.1 I will show that reason explanation of action has more force than causal explanation, which can be seen by looking at supported counterfactuals. Action explanation supports additional counterfactuals that are incompatible with causal explanation – therefore the claim that action explanations are causal hardly accounts for their force. That was, however, the core of **FPCE**'s claim. Moreover, I will suggest in section 5.1.4 that the counterfactuals which are supposed to show that action explanation is causal do not have the required empirical content.

2.2.8 Agent causation and natural causation

Some modern authors (for example, Roderick Chisholm and Richard Taylor) have argued that the concept of natural event causation cannot be used to clarify the concept of action, since the phenomenon involved in causation is a different sort of causation. This *sui generis* notion of causation has been termed agent causation, and the idea is that agents are, literally, the causes of their actions. As Richard Taylor[50] explains, agent causation is incompatible with the causal theory of action as we discussed it. For if the agent is the cause of his action, it cannot be the case that there was some event which was the cause of that action, since agents and events cannot be identical (and, presumably, as with the problem of mental causation, we want to rule out massive overdetermination of actions by their causes). At the same time we can see why agent causation is unacceptable to physicalists: if agents are admitted as causes, the causal network of physical

[50] Taylor 1966

causes and effects is no longer closed. But if these points are considered difficulties for agent causation, there are other considerations to be taken into account. It may be doing a better job at bringing out the special character of actions as opposed to other events. And given that it is the agent, rather than some physical event - even if it is instantiated by the agent - which does the causing, we can make more sense of the agent being the author of his actions.

A perhaps stronger objection to a causal theory of action has it that the concept of natural event causation is parasitic or dependent upon agent causation. In other words, the concept of causation as used in the natural sciences is really a derived anthropomorphic concept. It does need further argument to exclude the possibilities that the dependency runs the other way around, or that the two concepts of causation are somehow related but not in such a way that one is dependent upon the other. Richard Taylor's view[51] that there is a sui generis notion of agent causation existing alongside natural event causation is an example of the latter. I find such a view unattractive, for the simple reason that it is unclear why both these relations go under the same name of 'causation'. I will put this kind of view aside, and will now go on to consider the view that natural event causation is merely a derived notion.

The idea that event causation is parasitic upon agent causation is far from new. It was probably held by Locke[52] and certainly by Reid; a modern proponent of the view is von Wright[53]. The important issue dividing the ranks here, as we shall see, is: is the view that agents are the causes of their actions combined with some form of volitionism, according to which a volition exercised by the agent is the distinguishing mark of agency? Thomas Reid, e.g., held that only an intelligent being can exercise active power, and consequently it is only intelligent beings that can truly be called causes:

> "The name of a *cause* and of an *agent*, is properly given to that being only, which, by its active power, produces some change in itself, or in some other being. The change, whether it be of thought, of will, or of motion, is the *effect*. Active power, therefore, is a quality in the cause, which enables it to produce the effect. And the exertion of that active power in producing the effect, is called action, agency, efficiency.(...)
> It is very probable that the very conception or idea of active power, and of efficient causes, is derived from our voluntary exertions in producing effects; and that, if we were not conscious of such exertion, we should have no conception at all of a cause, or of active power, and consequently no conviction of the necessity of a cause of every change which we observe in nature. (Reid, essay iv ch.ii (p.603, 604))

[51] Taylor 1966
[52] Locke 1975
[53] von Wright 1971

The notion of causation cannot on such a view elucidate what action is, since action - i.e., an agent causing some change by having a volition - is itself the paradigm of causation.

The idea that natural event causation is parasitic upon action is not necessarily linked to a volitionist theory of action. There is no reason why we could not give some other account of action, and then proceed to argue that our notion of event causation rests upon that very notion of action. This is what von Wright does. The difference between Reid's and von Wright's position is that, for the first, event causation is parasitic upon the relation between an agent and his action, whereas for the second the relation between an agent's action and the result of that action is the primitive. I will not be concerned here with what von Wright says about action: for now we have to see what arguments he gives for thinking that action is the primitive, and event causation the derived notion.

von Wright says, in what I take to be an attempt at definition, the following about causation:

> "I now propose the following way of distinguishing between cause and effect by means of the notion of action: *p* is a cause relative to *q*, and *q* an effect relative to *p*, if and only if by doing *p* we could bring about *q* or by suppressing *p* we could remove *q* or prevent it from happening." (von Wright 1971, p.70)

The point can be expressed in terms of counterfactuals. If we want to establish that p caused q, then it is a necessary condition that the following counterfactual is true:

(CC) If p hadn't happened, q wouldn't have happened.

The crucial insight is that we can not establish the truth of this counterfactual by only the observation of p's and q's, because the most we will then see is a constant conjunction[54]. We therefore need to be able to experiment, i.e., to produce p at will at a number of similar occasions. In other words, we need to be confident that

(AC) If I had not produced p, p would not have happened.

Of this we are in practice quite confident: we feel that we can act. Not that we are sure of it - strange things can always happen - but on the whole we can rely on the truth of AC. If we could not, action would not be possible,

> "for it is an essential feature of action that changes should happen of which we can say confidently that they would not have happened had it not been for our interference, and also that changes fail to take place of which we

[54] Several assumptions are made here. Firstly, that the epistemology of the causal relation tells us something about what is conceptually primitive. Secondly, that constant conjunction is not the same as causation. Furthermore, in chapter 5 doubt will be cast on the assumption that a causal relation implies the truth of a counterfactual.

> can confidently say that they would have occurred had we not prevented them." (p. 61).

Why is it so important that we should be able to experiment and produce p at will? If we could not, we would not know whether the system under consideration is closed. If we say that p is the cause of q, that means that there is no other (feature of an) event besides p which is sufficient for q. That of course has to be put to the test, and we do so by reiterating a similar situation several times, producing or preventing p at will, and observing whether q indeed co-varies.

One may object that surely there are situations which cannot be reproduced experimentally, or that there are events which are outside our sphere of influence. Since this not stop us from making causal statements about such situations (nor indeed from understanding them), the objection would be, the experimentalist notion of causation must be wrong. For example, we say that the eruption of the Krakatoa caused a world-wide drop in temperature; or that the existence of a black hole causes that we do not observe any light coming from a particular place in the universe. However, this does not contradict anything we just said. Our understanding of such causal statements nonetheless rests on hypothetical experimental situations: *suppose* we could recreate the situation in which the Krakatoa erupted, and that we could in that situation prevent the eruption, would the temperature still drop world-wide? Nor is this simply a matter of wild speculation, since this kind of example concerns enormously complex causal relations, which can be broken down into simpler ones which often can be tested experimentally.

It is important to see that the action-counterfactual (AC) is not intended to be a causal counterfactual, and that on von Wright's view therefore agents are not the causes of their actions. They simply perform them: "..by *making* the cause *happen*, we achieve or bring about the same as the cause does by happening. To say that we cause effects is not to say that agents are causes. It means that we do things which then as causes produce effects, 'act' or 'operate' as causes." (p.69) For this reason it is inaccurate to say that he thinks that event-causation is parasitic upon agent-causation: his insight is better expressed by saying that event causation is parasitic upon action. Our understanding of what causation is and what causal statements mean rests upon our understanding of what it is to act. But whether or not von Wright's position amounts to agent causation - which depends on there being a real distinction between AC and CC - does not matter for present purposes: his arguments for event causality being a derived notion do not depend on that distinction. (What does depend on it is what exactly event

causality is derived from or dependent upon.) The result is the same: event causation cannot be of any help in elucidating the concept of action without creating a vicious circle.

Have von Wright's arguments conclusively established that event causation is parasitic on action, and not the other way around? Can one not argue plausibly that, instead, causation is the notion that underlies the idea of doing things, since it provides the basis for manipulating one thing by manipulating another? von Wright answers that to reason in this way would be to beg the question. The knowledge that q always succeeds p is not enough to provide a basis for manipulation: for that, the link needs to be nomic. In other words, if the constant succession we had observed were merely accidental, there would be no reason to think that we could effect q by doing p. The link between p and q must reflect some sort of necessity, which is expressed by CC; but what account is then given of how we understand CC? It must rest on AC, and so we have travelled full circle.

Name of position (proponent)	Classification as causal theory of mind	Objections ->	Infinite regress	Nature of cause: elucidatory?	Neutral bodily movements	Logical connection argument	Deviant causal chains	Epiphenomena-lism; mental causation	Rational actions	'Causal' explanation w/o causal relation	Causation conceptually dependent on action
Cause-volitionism (Jennifer Hornsby)	NCP		-	#	#?	*?	#	*	*?	-	*?
Hybrid volitionism (E.J. Lowe)	NCA		*?	#	#	*	#	*	*?	-	*?
Anomalous monism (Donald Davidson)	FPCE (FPCS, NCA)		-	#	#	*?	#	*	*	-	*?
Practical realism (L.R. Baker, William Child)	FPCE ¬(NCA, FPCS)		-	-	-	-	-	*?	-	#	*?

* constitutes an objection to the position;
\- does not affect position, or is satisfactorily answered
\# undermines an argument in favour, or motivation, of the position
? position not sufficiently worked out to assess impact of criticism, or controversial

2.3 Summary and conclusions

In the preceding sections I have attempted to draw a map of the various available causal theories of action, the objections to these, and the extent to which the latter affect the former. The results I have brought together in the above table, which is meant both as a reminder and overview of the discussions. As can be seen, none of the positions emerges unscathed. Prospects for the causal ancestry claim **NCA** don't look too good, whether defended on epistemological grounds (volitionism) or explanatory grounds (anomalous monism). The mere explanatory claim **FPCE** (practical realism) seems a bit less problematical.

At this point we can say a little more about the relationships between some of the main theses I set out in the previous chapter. If we combine **FPCE** with token-identity of mental and physical events, then we get **FPCS**. When it is not so combined, it may be questioned whether we have a coherent and intelligible position: which causal relations make the causal explanations of actions true, and if the answer to that is: none, then why are they causal explanations at all? If to **FPCE** + **FPCS** we add (as Donald Davidson does) the thesis that all actions are intentional under one description or another, and therefore that all actions are explainable by a (causal) reason-explanation, then we get **NCA**. **FPCS** forms the basis of the position known as functionalism in the philosophy of mind: the view that something is a mental state of type X if and only if it occupies a certain place in a causal network, and can therefore be characterised by typical perceptual inputs and behavioural outputs. **NCA** is stronger than that, since it lays down one specific necessary condition on causal ancestry, rather than vaguely indicate a typical causal network.

However, even if serious trouble has emerged for these causal claims, we are far from a 'blanket ban' on the use of causality in theories of mind. Firstly, it should be acknowledged that repairs might be made to the theories that I have considered. I hope, though, that causalists agree that the ball is firmly within their court. Objections have to be answered, or new arguments given to buttress their position. Secondly, other positions which one might call 'causalist', and which I formulated in the previous chapter, have not as yet been discussed, and (therefore) neither affected by any arguments given here. I shall come back to this point in chapter 5. Thirdly, we cannot prejudge from these discussions how the situation is with other concepts of mind. To remedy the last point, I will go on to discuss the causal theory of perception.

3 Perception

In this chapter I want to take a look at another mental concept about which causal claims abound in the literature. The chapter is structured in a way very similar to the previous one. First, I will look at general motivations for a causal theory of perception. Then I will describe some of the causalist positions in the literature, classifying them according to my scheme from chapter 1. I go on to discuss some important objections to causal theories of perception, and with an assessment of how the different types of causal theory of perception are affected by these objections.

3.1 Motivations for causal theories of perception

The first motivation is more like an intuition, and has to do with science. Anyone who knows a bit about physics and physiology is surely acquainted with the fact that objects reflect, of the light that falls on them, certain parts of the visible spectrum; that this light enters our eye through the pupil and is then projected by a variable lens on our retina; that receptors on the retina convert this light into tiny electric pulses travelling down our optic nerve to that part of the brain concerned with vision, the visual cortex. Given the unquestionable truth of these scientific facts, it is obviously true that in perception the object does causally affect the subject (it has to reflect light to it), and so a causal theory of perception is the obvious, nay, only, choice. E.J. Lowe, for example, writes: "It seems to me that any theory of perception that is to respect the known scientific facts of human physiology and the laws of physics must be a *causal* theory."[55] But here we have to be careful which claim exactly is being made. Only a claim of type **CPI** is available on the basis of this kind of argument, that is, a claim about what processes are as a matter of fact involved. A further claim about the *concept* of perception would need further argument. Consider, for example, whether we can make sense of a creature that is not like us causally affected by its environment, but nevertheless perceives things in that environment. If that is too far fetched, consider this: if we tell somebody who says to perceive an object that he is in fact not causally affected by that object, is that reason for him to retract his claim? In an example of Dretske's, a subject is placed on one side of a massive brick wall, and an object on the other side. The subject describes the object in detail; what's more, if the object is changed or moved he reports correctly on that. Naturally we are puzzled: obviously there is no causal link, so how can he perceive it?

[55] Lowe 1995, p.59

However, given this evidence, it seems hard to deny that he does perceive it; nor does this case, on the other hand, falsify the empirical claims about what is typically happening in cases of perception.

Incidentally, not only do we need to take care with the modal status of claims made on the basis of this motivation; it is also important to consider what exactly the effect-end of the involved causal link is supposed to be. Most causal theories of perception, as we shall see, make some kind of claim about this, but such claims are very rarely supported by scientific evidence.

A common theme in much of the literature on causal theories of perception has to do with the phenomena of illusion and hallucination. The aim, in such discussions, is to achieve a philosophical understanding of the grounds on which the distinction between illusion, or hallucination, and genuine visual perception is based. Especially in the case of hallucination it is natural to think that what distinguishes it from genuine visual perception must be something about the relation between the perceiver (subject) and the perceived (object). The claim that the salient difference is the presence or absence of a suitable causal relation is then an obvious candidate. Claims made in these discussions are typically of the type **NCA**, but the bolder claim **NSCA** is sometimes not far off. (Two varieties of such claims can be distinguished here, depending on what they are supposed to be stating conditions for. The relevant distinctions here can be either between seeing and hallucinating, or between seeing one object rather than a (similar) other one. If only the latter claim is made, then arguably it is of the type **VCD** ('certain valid distinctions among mental concepts are causally-based distinctions')) We will also see that the weaker **FPCE**-type claim is made by at least one author, that is, a claim to the effect that explanations of the occurrence of visual perception are causal, rather than a claim about what conditions of application of the concept are.

The third motivation is concerned with epistemological considerations: how can we come to have knowledge about the world? Knowledge of the world is gained, obviously, by perception. And for that knowledge to be true knowledge, it needs to be reliable: if it were a hit-and-miss affair, we would hesitate to call it knowledge. Perception, the means by which the knowledge was gained, thus needs to be reliable as well. In order for this to be so, the states of affairs that we perceive have to be actively involved in, i.e. causing, our perception of them. Claims in this area can be of the type **NCA** or **FPCE**, as we shall see.

3.2 Some causal theories of perception

3.2.1 Sense-data causalism

The causal theory of perception as it figures in the works of amongst others Grice[56], Pears, and Strawson, is formulated by Snowdon[57] as a commitment to the following three claims:

(1) It is necessarily true that if a subject S sees a public object O then O causally affects S.
(2) O must produce in S a state reportable in a sentence beginning 'It looks to S as if ..'
(3) Theses (1) and (2) represent requirements of our ordinary concept of vision.

I will hereafter refer to these three claims together as **SDC**.
Condition (1) is very broad; it is easily fulfilled by many objects that S doesn't see. For example, α-radiation will affect S if he is exposed to it, but nonetheless it is invisible. A heavy object falling on his head will affect S, although he doesn't see it. Something, therefore, needs to be said about the sort of effect that O has on S, and this is what (2) does. Condition (2) is neutral between subject S having a hallucination of an O and S genuinely perceiving O. A perhaps more common way of putting this condition is that O must produce in S a visual experience as-of an O. That formulation has been taken by many to introduce extra, and unwanted, items into our ontology, namely visual experiences. It needs argument to show that the above formulation does *not* do so: of what character is the 'state'? In following sections this will be discussed extensively.
It is important that both (1) and (2) are claims stating necessary conditions for application of the concept of vision (or visual perception). Conditions are given, that is, which need to be fulfilled if S is to genuinely see O, but which do not guarantee that such will be the case; namely, (1*) O must causally affect S, and (2*) the effect of O's causally affecting S must be a state of a certain sort. From this it can be seen that this version of the causal theory of perception is, in the terminology of chapter 1, an **NCA**-type claim.

[56] Grice 1961
[57] Snowdon 1981

Condition (3) just says that (1) and (2) are conceptual truths, rather than for example empirical facts. This is meant to explicitly state commitment to more than the **CPI** claim.

3.2.1.1 The Argument from Illusion

The three theses **SDC** are usually defended by means of the so-called argument from illusion, which proceeds as follows. A mere correspondence between what a person takes himself to see, and what is there before his eyes, does not by itself establish that what we have is a case of genuine perception. The correspondence may be a coincidence; or it may have been contrived, rather than resulting from the normal exercise of his visual faculty.

For example, suppose we seem to see a pillar over there. Indeed, there is a pillar over there; but what we don't know is that between us and the pillar there is a mirror, reflecting the image of a numerically different, though similar pillar[58]. Pears' example, which involves hallucination rather than illusion, is that of the thirsty traveller in the desert, who has a mirage of an oasis: as it happens, he has indeed reached an oasis similar to the one he hallucinates.

In both these cases we do have a correspondence between what the subject takes himself to see, and the scene in front of him. Still, we would rightly hesitate to call these cases of genuine visual perception. But we have some idea about what is missing in each of these cases. What's wrong is that the objects before the subject play no role in the visual experience had by him. Genuine vision, on the other hand, should inform us about how things are in the world. Therefore, in cases where we can correctly say that someone sees something, there should be some connection rather than just correspondence, and a causal connection seems to be the obvious candidate.

As J.L. Austin[59] has pointed out, illusion and delusion are not the same thing. Therefore, to avoid muddling the argument, I will from now on by default only speak of cases of delusion (of which hallucination is a special case) as opposed to cases of genuine perception.

The Argument from Delusion (**AFD**) could be formalised as follows:

(A1) It is logically possible that genuine perception and delusory perception are indistinguishable to the subject.

(A2) The visual experiences involved in genuine perception and delusory perception are in such cases indistinguishable to the subject.

---

[58] Grice 1961, p.69-70

[59] Austin 1962 p22 ff.

(A3) The visual experiences involved in genuine perception and delusory perception must be the same in such cases.
(A4) The difference between perception and delusion lies therefore in some extrinsic property of the visual experience involved.

This argument is cautiously formulated in that it speaks of a logical possibility, rather than asserting that such situations actually happen. It may be that in all the experiments we conduct, the subject always finds a way of telling whether he is hallucinating or genuinely perceiving. Should that happen, though, it would be no problem for **AFD**: the only thing that we need to ask is an explanation of a logical possibility.
Note that the argument is not conclusive, because (A3) fails to follow from (A2). Things are not necessarily the same, just because we cannot distinguish between them. Most people cannot distinguish between female and male chicks; yet obviously there is a difference, and chicken-sexers can see it. There is, however, something awkward about treating the two kinds of visual experience as very different. If we fail to distinguish two visual experiences from one another, what other criteria for individuation could there be? We feel that, at least, we have to be able to give an explanation of (A2); and (A3) seems to be the best explanation. This inference to the best explanation is what is attacked by proponents of the so-called disjunctive conception of visual experience, to which I will come back shortly.
Another important point to note is that **AFD** is silent about what the extrinsic factor in (A4) is supposed to be. In order to arrive at the causal theory of perception (**SDC**), it needs argument to show that the extrinsic property in question is the property of being caused by the object of the visual experience. Therefore, **AFD** needs to be supplemented; this can be done by an argument from counterfactuals. The idea would be that in cases of genuine perception, certain counterfactuals are true, such as "Had there not been a platypus in front of Pia, then Pia would not have had a visual experience of a platypus". What makes such counterfactuals true, the argument continues, is that the platypus' presence causes Pia's experience of it. In this way, **AFD** may be supplemented. It may in fact be argued that this argument can stand on its own, and does not need **AFD**. Thus it might[60] form an argument for a causal theory of perception without any commitment to visual experiences, which as we shall see, leads to trouble. I defer

[60] In an adapted form, namely by changing the consequent of the counterfactual to "..., then Pia would not have seen it".

explicit discussion of this self-standing argument, which has a close analogue in action, to chapter 5.

3.2.2 Disjunctive causalism

3.2.2.1 The disjunctive conception of experience

The **AFD** requires a specific conception of experience, which could be termed non-disjunctive. The argument said (A3) that the visual experiences involved in delusion and genuine perception are the same. The visual experience is thought of as a basic world-independent item, in terms of which an account can be given of perception that enables us to distinguish it from delusion - namely, by postulating that in cases of genuine perceptual experience there is some extrinsic feature not present in cases of delusory experience.

A common objection to such inferential or indirect realist theories of perception is that they lead to scepticism about our knowledge of the external world. If our perception of the world is never direct but always via some epistemic intermediary, whether we call it a visual experience or a sense-datum, then we can never be sure whether what our senses tell us is true. Indeed, we cannot even be sure whether the material world exists at all - we could be consistently deceived by a Cartesian evil demon. The best an indirect realist can come up with is that the existence of the material world is an inference to the best explanation: Russell[61], for example, tells us that although it is possible that there be no material world, we have no reason to believe this to be the case. But that leaves the sceptical thought that we do not know that it is not the case unaffected. Such scepticism is unacceptable; therefore, we want a theory of perception which allows that we are directly aware of the world, without intermediary. This can be delivered by a disjunctive conception of experience: having an experience, according to it, is either a case of direct awareness of facts in the world, or (in the case of delusion) seeming to see that which is not true. This is how McDowell[62] presents the matter.

Another motivation for proposing a disjunctive conception of experience is simply centred on resisting the **AFD**. The conclusion (A4) that some extrinsic property of the visual experience involved must make the difference between delusion and genuine perception is of course escaped by countering that 'the visual experience' was not one and the same thing in the two cases to start with. It

[61] Russell 1967
[62] McDowell 1982

just so happens that the concept of a visual experience applies to different cases, but that by itself does not mean that we have a common factor. Perhaps this approach may strike one as rather ad-hoc, and not a good argument for holding the disjunctive theory to be true. Note, however, (as Snowdon 1981 does), that the disjunctive theory *need not be true* to be a counterargument to **SDC**. It is sufficient that it *not be conceptually false*, i.e. that it be merely possible that the concept of experience is disjunctive in character. This is due to condition (3) of the **SDC**, which says that it is a conceptual truth that the subject is causally affected by the object in cases of genuine perception. If we can imagine that the concept of experience is disjunctive in character, then we can imagine the possibility that the conclusion of the **AFD** does not follow: and thus thesis (3) of the **SDC** stands unargued for.

The disjunctive theory reverses the order of explanation[63]: visual experiences are not explanatory of what perception and delusion are, but the other way around. A visual experience is what the subject has in *either* a case of delusion, *or* a case of veridical perception. The concept of experience, in other words, is disjunctive. The disjunctive view of experience paints a picture in which (A3) doesn't follow from (A2), and thus shows by example that AFD is not cogent.

3.2.2.2 Disjunctive causalism

Above it was suggested, following Snowdon, that the disjunctive conception of experience hands us a way to resist **SDC**. But is there a way to formulate a causal theory of perception that is compatible with the disjunctive conception of experience? Child[64] thinks this to be the case. He concurs with McDowell in endorsing the disjunctive conception of experience, on the grounds that

> "on the non-disjunctive conception of experience we are not in direct cognitive contact with the world, since the most basic characterisation of experience is world-independent. But it is arguable that no concept constructed solely from world-independent contents can itself be a concept of an objective world independent of thought: if that is right, then no theory, or inference to the best explanation, could get us from experience conceived as a highest common factor to thought about the world." (Child 1994, p.149)

This is an argument about the epistemical predicament we are in, based on conceptual considerations. As a consequence, no empirical evidence can provide a counterargument to the disjunctive conception. Suppose, for example, that neurophysiologists found that there was a physical state in the brain in common

[63] see Child 1994
[64] Child 1994 p.194 ff.

between a case of genuine perception and its delusory counterpart. Robinson, attacking the disjunctive theory in order to argue for sense-data, says about this:

> "It is necessary to give the same account of both hallucinating and perceptual experience when they have the same neural cause. Thus, it is not, for example, plausible to say that the hallucinatory experience involves a mental image or sense-datum, but that the perception does not, if the two have the same proximate - that is, neural - cause." (Robinson 1994 p.151)

But, contrary to Robinson, such a discovery could not establish that the disjunctive theorist is wrong: for he can still allow that there are physical / causal intermediaries in common between perception and delusion, as long as the subject has no epistemic access to such states. In other words, such merely causal intermediaries would perhaps be sub-personal informational states, but no epistemic intermediaries. The empirical discovery would thus have no grip on disjunctivism.

It is important to see that disjunctive causalism is quite different from **SDC**. That theory explicitly committed itself to the existence of a causal intermediary to which the subject has epistemic access, in its second thesis:

(2) O must produce in S a state reportable in a sentence beginning 'It looks to S as if ..'

William Child has surreptitiously[65] changed this to:

(2') O is causally responsible for a state of affairs reportable by a sentence of the form 'It looks to S as if...'

The important difference is that there is no longer any mention of a state *produced in S*. Instead, there is a 'state of affairs', which may, as Child later explains[66], be interpreted as a relational state, making irreducible reference to O. In other words, this state of affairs can simply not occur if O is not there. There is therefore no commitment to the existence of a world-independent internal state which the disjunctivist worried would lead to scepticism. I will refer to this position – **SDC** with condition 2 replaced by 2' – as **DISC**.

A second important difference is that the vocabulary has shifted from "O must [causally] produce..." to "O is causally responsible for...". It would be odd to say that O causally produces a state of affairs of which its own existence is a part,

[65] Child 1994, p.141. 'Surreptitiously', because he remarks in a footnote that his formulation is 'adapted from Snowdon', but does not spell out what the important difference is.
[66] Ibid. p.161

and retaining the old vocabulary as well as accommodating the disjunctive conception of experience would require doing so. Now it might be argued that this is no problem: it is not obvious, for example, that we cannot say that the bombing of Pearl Harbor caused the American involvement in WW II, of which it was at the same time a part. But Child chooses to be non-committal about which states cause which other states, and prefers instead saying that we need a causal-explanatory notion; 'causally responsible' is intended as such. It is unclear whether the disjunctive causalist theory of perception should, according to my classification, be classified as an **FPCE**-type theory or an **NCA**-type theory. On the one hand, Child distances himself from talk of one state causally producing another, but prefers causal-explanatory language instead. But on the other hand, the claim he defends is hardly about the character of explanations that 'the folk' give of perceptions – do we give such explanations at all? – but rather about conditions of application of the concept of perception. It is not my classification which is at fault here, but a tension in Child's work. His argument is set up in such a way that the causal theories of action and perception that he defends are similar; but as I have shown they are not.

Note also that the **AFD** cannot figure as an argument for a disjunctive causal theory of perception. The argument which many take the **AFD** to provide says that since there is a common factor, there must be some differentiating factor extrinsic to the experience. But if we deny, with the disjunctivist, that there is such a common factor, then obviously there is no pressing need for a differentiating factor - since a delusory experience is already different from a perceptual experience, we don't need to heap yet another differentiating factor on top of that. The causal theorist, then, can be a disjunctivist, but only at the price of giving up his strongest argument. Instead he could rely, as Child does, on the argument that perception needs to be reliable, perhaps buttressed by the argument from counterfactuals to which I have alluded earlier. The reliability argument will be further discussed in section 3.3.4, and the argument from counterfactuals in chapter 5.

3.2.3 Experientialism

Many philosophers take the disjunctive conception of experience to be a rather desperate move, inspired by reading too much into talk of indistinguishable experiences. What puzzles them is why one would want to say that in both the hallucinatory and the perceptual case it looks to S as if an O is there, but in the same breath deny that S has the same visual experience in the two cases.

"Surely", they say, "it is against common sense to deny that genuine perception involves the same visual experience as its delusory counterpart. But such a common factor does not commit us to indirect realism and sense-data."

According to the indirect realist, if a rose appears pink to me but is in reality red, the object that I perceive cannot be identical with the object that is there in front of me (viz. the rose). For that would require that one object be red and pink at the same time in the same place; and so it is concluded that what we are directly aware of is an intermediary object or sense datum. However, an objection to this reasoning is that a fallacious inference has been made; we cannot infer from 'X appears pink to me' that it must be the case that 'there is a pink X that appears to me'. It is indeed true that from 'John appears angry to me' it doesn't follow that 'An angry John appears to me.'[67]. I will not follow Robinson here in discussing at length whether what he calls the phenomenal principle ("If there sensibly appears to a subject to be something which possesses a particular sensible quality then there is something of which the subject is aware which does possess that sensible quality."[68]) is true. Whether it is true or not is not really at issue here: I will, guided by a representative recent example, try to show that if visual experiences are regarded as explanatory items in the **AFD**, then the truth of the phenomenal principle is assumed, and sense data are inescapable.

The sceptical argument, denying us knowledge of the material world, only has grip if visual experiences are things that the subject must be aware of. Only on such a picture are visual experiences epistemic intermediaries, and only on such a picture is a vicious regress started off by the question: how, then, is the subject aware of these visual experiences? However, this is not the right way of thinking about visual experiences: the subject is not aware of them, he just has them (although that is a fact that he then can be aware of).

> "The common element [...] is the having of a visual experience such that it looks to the subject as if a dagger is there, but such an experience is not an object of perception nor are its features objects of perception. The experience, rather, is something the having of which would, if certain other conditions were satisfied, count as perceiving a dagger." (Millar 1995, p.82).

On this view, which Millar calls the experientialist view, visual experiences are a type of psychological state:

> "According to the experientialist the subject of a genuine perception and the subject of a hallucinatory counterpart of such a perception will have in common a certain psychological state, an experience of a certain type the having of which does not necessarily depend on the presence of the object perceived in the perceptual case." (Millar 1995, p.75).

---

[67] see Robinson 1994

[68] Robinson 1994, p.32

There are, however, problems with such a view. People *have* visual experiences but they *are in* psychological states. This may seem an unimportant quibble; but given the different uses of the two terms, can it make sense to say that to the subject two psychological states are indistinguishable? What I am aware of is having certain visual experiences, and being in certain psychological states. Perhaps, and I think that this is a plausible construction of what Millar wants to say, these are the indistinguishabilia. But on the **AFD**, it is the visual experiences that are indistinguishable, not the havings of visual experiences. This is a different level of indistinguishability; it weakens the **AFD**, for we could imagine two different visual experiences the having of which is indistinguishable. The common element becomes less inescapable.

A related point is that if it makes sense to say that visual experiences are indistinguishable from one another (to the subject), then surely they must be items of awareness (to that subject). This may seem to be cheating, on the grounds that the term 'distinguishing' is also used in many other contexts: e.g., we distinguish between letters and numbers, between axioms and premises, between animate and inanimate objects, etc. The visual sense of distinguishing is not the only one. However, this does not help much: for in this logical sense, we can, and do, distinguish perceptual visual experiences from delusory visual experiences. The point was not that in general we can't tell whether a subject is genuinely perceiving or not: usually we are in a position to know whether the object purportedly seen is actually there, or whether otherwise things are as the subject reports them. But the sense in which the subject can't distinguish between the two must be the visual sense. If we tell a subject who perfectly well knows the difference between hallucination and genuine perception that the object he reports to be seeing is not there, it still makes sense for him to say, in the sense relevant to the **AFD**, that the visual experience he has is indistinguishable from one he might have in genuine perception, even though in the logical sense he does make the distinction (for he takes it on our word that the object is not there.) In conclusion: if it is denied that visual experiences are items of awareness, then it doesn't make sense to say that they are (in)distinguishable.

I conclude that the experientialist theory of perception is either no different from sense-data causalism, or does not have the resources that it pretends to have in order to deal with the argument from delusion. In any case, given that it is meant as a sophistication of **SDC**, it is – according to my classification – an **NCA**-type theory. The same thing is true of the following two theories that I will consider.

3.2.4 Adverbial theory

Another theory eager to reject the 'sense-datum fallacy' is the adverbial theory. (see Chisholm 1966. A modern form of this theory, which avoids a number of objections to the earlier versions, is Tye's operator theory[69]). The thought driving this account is that the surface-grammar of perception-statements is misleading. We shouldn't think that

(1) Michael drives a car

has the same structure as

(2) Michael sees a car

We should rather see the latter on a par with

(3) Michael gave an impressive argument

which can be recast as

(4) Michael argued impressively.

Similarly, we should think of (2) as transformable into

(5) Michael (visually) senses (a-car)-ly

In other words, what we usually call the object or content of perception should be seen as an operator modifying the basic form of (in this case, visual) perception. Grammatically, this can be expressed by using adverbs.

The claim of adverbial theorists is that they can account for the phenomenology of perception, while keeping ontological commitments to a minimum; there is no need for quantifying over sense-data. Looking at (5), this seems to be right: there is no commitment to the existence of a real material car in front of the subject, nor to the existence of a sensum of a car. The subject is just sensing, and does so in a specific manner (which leads him to say, e.g., that there is a car in front of him). However, (2) does imply that there be a car in front of Michael, which means that (2) and (5) are not equivalent. (5), then, must be a translation of:

(6) Michael has a visual experience as-of a car.

(6), like (5), is neutral as to whether there is a car, but (5) is thought to be a better formulation since it seems to avoid the introduction of experiences - whatever they are - in our ontology.

Has the adverbial theorist answered the **AFD**'s challenge? Does (5) explain why Michael might mistake himself to be perceiving when in fact he is hallucinating? That seems to be open to dispute. (5) does *describe* a situation which is neutral as between perceiving and hallucinating, but I don't think it can have the pretence of *explaining why* that situation has this neutrality. If we compare (5) with (6), the pretence of giving an explanation seems to have been dropped: what is left is a statement of brute fact, 'What perceiving that F has in common with

[69] Tye 1993

hallucinating that F is that in both cases one senses F-ly.' The indistinguishability between the two situations to the subject has as it were been built in, without trying to bother to explain why there is this indistinguishability. Now, as I will argue later on, it is a mistake to accept the **AFD**'s challenge. But what we need to determine is: does the adverbialist take up the challenge? The adverbialist merchandise is mostly praised by saying "It accounts for the phenomenology of perception as well as a sense data theory does, but without the postulation of dubious inner items", or "It can do what the sense-data theory does, but with greater ontological simplicity". If we are to take these slogans at face value, they must imply that the adverbial theory can confront the **AFD** like the sense-data theory can. But the adverbial theory has only the mere pretence of offering an explanation, without delivering the goods. If the adverbialist wants to distance himself from this one task that the sense-data theory has taken upon itself, he'd better say so.

3.2.5 Dispositional (belief-)theory

A causal theory related to the experientialist theory in its effort to eliminate intermediary items of awareness is the belief-theory, found in Armstrong 1968 and Smith&Jones 1986. The idea of this theory is that the effects caused in the subject by the object are not visual experiences, but perceptual beliefs. Smith & Jones put it thus:

> "Perception, we noted, is the major route by which we acquire information; it is the primary means by which we come to form new beliefs about how things stand with the world. So perhaps we should say the end effect of the perceptual process is exactly that - the acquisition of beliefs. In other words, what is essential to perception is not the receiving of, say, pictorial representations before the mind's eye but the receiving of information." (Smith and Jones 1986, p.104)

(They go on to argue that, since the information-transmitting process needs to be reliable, there needs to be a causal link. This argument I will discuss later.) The proposal will not do as it stands: sometimes one has overriding reasons not to believe what one sees. Perhaps, then, we should speak of a disposition to form perceptual belief. However, Jackson[70] points out that as long as we don't have a non-circular answer to the question of which conditions have to be met for the disposition to actualise, we get no further. If we give a counterfactual account of the disposition, as Armstrong chooses to, how do we decide which is the relevant counterfactual?

[70] Jackson 1977

> "For there will, of course, be sentences 'p', 'q', such that, in the case of the wall, 'If p, then I would believe that the wall were blue' and 'If q, then I would believe that the wall were white'; and the wall does not (and cannot) look both blue and white.[..].I think the most that can confidently be claimed in the way of relevant counterfactuals in the case where the wall looks blue, is something like: if I had not known or been fairly certain that the wall was white and if I had believed that the circumstances were such that objects look the colour they are, then I would have believed that the wall was blue. But such a counterfactual achieves nothing in the way of a belief analysis of 'looks', for it itself contains 'looks'..." (Jackson 1977, pp.40-41).

Apart from Jackson's objection, it is not clear to me that the belief-theory is not an indirect realist theory. Are the (dispositions to form) perceptual beliefs not introduced so that the description of the subject is independent of how things are in the world - and doesn't that make them into epistemic intermediaries, of the sort that I have shown leads to a problem of scepticism?

3.3 Objections to causal theories of perception

3.3.1 Conjunctivism and disjunctivism

3.3.1.1 'Indistinguishability': commitments of the disjunctive theory

I will now consider how we might understand the disjunctive theory; I will argue that the prevalent interpretation[71] of it gets into trouble by accepting that (A2) does stand in need of explanation. It may seem that the disjunctive conception of experience has robbed itself of the devices needed to explain why a subject may mistake delusory perception for genuine perception. The disjunctive strategy is to reverse the order of explanation: it is not the fact that delusion and perception both involve visual experiences which explains their indistinguishability, but the indistinguishability which explains why what is involved in delusion and perception are both visual experiences, albeit not the same visual experience in each case. Obviously, if the indistinguishability explains something about the visual experiences, the converse cannot also be the case on pain of circularity.

The case is not as clear-cut as I have presented it in the previous paragraph, because there is an equivocation in speaking of indistinguishability. It is crucial that we be clear about the question, 'indistinguishability of *what*?'. More specifically, we need to be clear about whether we are speaking of indistinguishability of cases of delusion and perception, or indistinguishability of visual experiences involved therein. To give an example, two cars can be

[71] as found in Snowdon 1981 and 1990, and McDowell 1982

similar[72] because they are both of the same model, or because they are both of a similar model. The model could be similar without being the same. The fact that the cars are similar explains that they must both be of some model (as opposed to being a shapeless, amorphous blob; hardly ever the case with cars). But the fact that the *models* are similar explains how the *cars* can be similar (in different cases, similarity of cars might be explained by e.g. their being made of the same material, or having the same wheels). Let 'genuine perception' stand for 'car A', 'delusory perception' for 'car B', 'perceptual experience' for 'model I', and 'delusory experience' for 'model II'. The non-disjunctivist says that car A and car B are similar, because model I (car A) is the same as model II (car B); the disjunctivist says that they are similar because model I is similar to model II. So, on the disjunctivist model, the similarity of car A to car B explains that they are both of some model: but the similarity of the models explains the similarity of the cars. These are all good explanations.

However, the disjunctivist does get into trouble by admitting that the visual experiences in the two cases are similar, if not the same. Similarity is a two-place (or many-place) relation. Talk about similarity, therefore, requires at least two similars, which must be the same kind of thing. By this I do not mean that a real apple cannot be similar to a cleverly crafted wax model: I mean that an apple cannot resemble a mental state, any more than a proposition can resemble a cat sitting on a mat. Perhaps mental states can represent apples (and propositions be about cats); but they do not do so by resembling. A mental state cannot be round and red and juicy, so we won't mistake it for an apple - just like a picture of a car doesn't drive you anywhere or consume any petrol. Now suppose that I hallucinate a horse in my garden: there is, in this case, no public object of which we can say that it is similar to a real horse. *It must therefore be an object of inner awareness - a sense datum or sensum.* The usual move here is to say that both visual experiences are 'as of a horse'; the contents are the same. To this the disjunctivist objects that this is a common-factor approach; we cannot say that even in the veridical case my experience is not of a horse, but only as-of-a-horse. However, this makes it impossible for the disjunctivist to explain the similarity between an experience of-a-horse and an experience as-of-a-horse. The answer that in both cases the same concept is employed ('horse') will not do, because this presupposes a picture in which the subject exercises his conceptual capacities on some sort of Given, or sense-datum.

[72] I switch here from 'indistinguishability' to 'similarity'. I can legitimately do so, because the first implies the second. My argument is intended to show that the weaker notion of similarity has undesirable implications for the disjunctivist.

The acceptance of (A2), therefore, forces a commitment to sensa. Since the disjunctive theorist disputes that (A3) follows from (A2), but does not dispute (A2) itself, he is in trouble. The only difference between the causalist and the disjunctivist story is that the sensum needn't be, indeed isn't, *one and the same* object in cases of delusion and perception.

If this is right, then a disjunctivist like McDowell is committed to sensa in spite of himself. His main argument for the disjunctive theory of experience is that in perception there is an unmediated openness of the subject to the facts; what is given to our experience in cases of perception is "the fact itself made manifest". McDowell apparently thinks the AFD's challenge can be met:

> "The most obvious attraction is the phenomenological argument: the occurrence of deceptive cases experientially indistinguishable from non-deceptive cases. But this is easily accommodated by the essentially disjunctive conception of appearances that constitutes the alternative. The alternative conception can allow what is given to experience in the two sorts of case to be the same *in so far as* it is an appearance that things are thus and so; that leaves it open that whereas in one kind of case what is given to experience is a mere appearance, in the other it is the fact itself made manifest." (McDowell 1982, p.214)

McDowell seems to say here that a 'mere appearance' can be similar to, or even indistinguishable from, a 'fact itself made manifest'. But if I am right, he omits to reflect on what sorts of things can be similar to one another. If there is some sort of sensum in the delusory case, as McDowell seems to admit, there must be one in the genuinely perceptual case too; otherwise the indistinguishability relation doesn't have the proper relata.

3.3.1.2 Conjunctivist and disjunctivist 'explanation'

Let us try to get clear what exactly is the demand for explanation. (A2) stated that "The visual experiences involved in genuine perception and delusory perception are sometimes indistinguishable to the subject." The question, then, is: how can this be so? But *do* we need to explain why a subject may be inclined to confuse delusory experience and perceptual experience? I suggest we take a critical look at the so-called explanation the causal theorist offers us. "Delusion and perception are indistinguishable because they are both as-of-O (the object); and this is because they involve the same visual experience." This explanation is modelled on something like this: "This sphere and that cube are indistinguishable because they are both yellow; and this is because they have been painted with the same paint.". The first part may look like an explanation, but isn't: we are just told *in which respect* two things are indistinguishable, so that we know what to

pay attention to. ("My car and this computer are indistinguishable because they are both made in Japan.") Similarity is never brute; it is always similarity in-one-respect-or-other. The second part does purport to give an explanation of *why* this similarity obtains; everybody understands, and will therefore accept as an explanation, that if you paint two objects with the same paint then (normally) they will have the same colour afterwards. In the case of visual experiences, however, the 'explanation' is different. I already pointed out that the locution 'as-of-O' is not very helpful for stating in what respect similarity obtains. As for the second part: do we *explain why* two things are indistinguishable by saying that they both fall under the same category? Such 'explanations' follow the model of:

> (**CE**) x and y are indistinguishable [in respect R] because x and y are both Ks.

But what about the following 'explanation'?

> (**DE**) x and y are both Ks because x and y are indistinguishable [in respect R].

(The latter, **DE**, is what Child thinks the disjunctive theorist offers us: he 'reverses' the traditional explanation.)

CE and **DE** , of course, do not explain the same thing. Explanans and explanandum have been swapped. But surely, on pain of circularity, they can't *both* be good explanations, so **CE** or **DE** - or both - have to be rejected. In effect, we are talking about the problem of Universals here. **CE** is the platonic 'solution', criticised by Wittgenstein[73] in his discussions on family resemblances, on the grounds that an essentialist definition of K is usually not available. Bambrough[74] then declared that Wittgenstein had solved the problem of Universals by reversing **CE** into **DE**- a dream dispelled by Manser[75]. This is, of course, the history of a huge debate in a tiny nutshell; but if we learnt our lessons from it, we should reject both **CE** and **DE** on Manser's grounds: explanans and explanandum are not independent. We don't know that something is a K independently of knowing in which respect it is indistinguishable to other Ks, nor the other way around.

So my criticism is that the causal theorist does not give the explanation that he thinks he does, for what he does is at most to provide a statement about the respect in which there is similarity. Disjunctivists who reverse the 'explanation' don't do any better, for the simple reason that explanans and explanandum are not independent. It is just inherent in the concepts of delusion and perception that we may be confused as to whether we see something or e.g. hallucinate it: and this is

[73] Wittgenstein 1953
[74] Bambrough 1960
[75] Manser 1967

brought out, illuminatingly or not, by saying that both involve visual experiences with indistinguishable contents. That both delusory and genuine perception involve visual experiences should however not be seen as an explanation of how they can have indistinguishable contents. The experience is what has the content: talk of visual experiences, therefore, just aims to clarify further what we mean by indistinguishable content - it is experiential content. The further question: why do delusory and genuine perceptions involve visual experiences with indistinguishable content?, is simply not meaningful. That is just how we use the concepts of delusion and perception. Why is it that chairs and tables are indistinguishable, in that they are both furniture? To answer this is to solve the problem of Universals. Of course neurophysiology may explain why we have such things as delusory experiences (and, for that matter, perceptual experiences), just as we can explain why there are chairs and tables: they don't grow on trees but are made by the carpenter. But being of a certain neurophysiological character does not turn an occurrence into a visual experience, any more than having been made by the carpenter turns chairs and tables into furniture.

To summarise: the sense-data causalist does not answer the **AFD**'s challenge. Nor does the disjunctivist who reverses the 'explanation' that the causalist gives. The explanation is just not to be had, and the challenge must be rejected. Looking back at the **AFD**, we can agree that genuine perception and delusory perception are sometimes indistinguishable to the subject (A1), and perhaps also that this can be re-phrased as saying that the visual experiences involved in the two cases are indistinguishable (A2), as long as we don't think of (A2) as somehow constituting or underlying (A1), thereby making visual experiences into explanatory items which could illuminate how (A1) can be the case. It should be noted that the disjunctive theory of experience, in its bare form which states that the truth condition for "S has a visual experience" is disjunctive, is compatible with these insights. Indeed, Hinton 1967, the original source of the disjunctive theory is plausibly construed as not making any explanatory commitments: these are emerging only later in the (more influential) papers of Snowdon and McDowell.

I have construed the discussion about the **AFD** as being about explanation: can the appeal to indistinguishable visual experiences explain why a subject may mistake delusion for genuine perception? Any theory which seeks to minimise its ontological commitment while endorsing the **AFD** paints itself into a corner, for it is the explanatory requirement made implicitly in the **AFD** which is the root of the problem. It is not so much a witch-hunt of sense-data that interested me: my point is that any theory denying the existence of sense-data (items of awareness,

visual experiences) cannot meet the challenge of the **AFD** to explain why a subject might mistake delusion for perception. Any theory which does commit itself to sense-data cannot do so either, and moreover leads to scepticism. **AFD**'s challenge is thus not met, and should, I suggest, be rejected.

3.3.1.3 Veridical hallucinations

There may be a residual worry about rejecting the argument from delusion and all that comes with it. It is all good and well to say that the challenge to explain why a subject may mistakenly think he perceives when he in fact hallucinates needs to be rejected; but how are we at all to make a distinction between veridical hallucination and genuine perception? That is to say, have we not given all our tools for making this legitimate distinction away, in rejecting the common-factor-plus-extrinsic-link account of hallucination and perception?

The causal account did not only attempt (unsuccessfully, as we have seen) to explain the possibility of a subject's mistake, it also enabled us to make the distinction between hallucination and perception in the first place. In most cases that distinction is of course very easily made: if somebody complains to me about the platypus marching about in his room while none is there, then the lack of fit between the world and his description tells me that he hallucinates the platypus instead of perceiving it. But this cannot be the whole story, since it seems that we can at least conceive of a situation in which what a subject describes himself as seeing perfectly matches the world around him, but in which nonetheless he falls short of genuinely perceiving these things. If this is indeed a possible situation, then how else can we make the distinction but by recourse to the absence or presence of a causal link between perceiver and perceived?

Wilkie[76] points out that cases of veridical hallucination play a central role in the causal theory of perception: "..veridical hallucinations were supposed to have a special role, rendering explicit a theoretical lynchpin of common-sense or ordinary language"[77]. This is so because the causal theory of perception has a dual task: "..demonstrating the presence of the causal condition in the ordinary notion while, on the other hand, explaining why it is (and largely remains) hidden in everyday discourse."[78] However, as a result, the causalist's arguments suffer from "a lack of friction with everyday situations"[79].

[76] Wilkie 1996
[77] ibid. p.251
[78] ibid. p.246
[79] ibid. p247

When talking about veridical hallucinations we need to be careful about what the envisaged situation is exactly. We should be mindful of the - perhaps not particularly funny - riddle: "It looks like a banana, smells like a banana, feels like a banana, and tastes like one. What is it? ... A banana!" Should the subject's description of the world in all details perfectly match the description of a genuinely perceiving subject, it would seem perverse to deny that the first is not, after all, genuinely seeing. Yet this is supposedly the kind of case that we are to consider, because if there were any mismatch then we can cry "hallucination!" on that basis. So if we want to deny that the first subject is genuinely seeing, we must have another reason for that than a mismatch. Suppose that the difference is that we know there to be no causal link in the case where we want to deny genuine perception. I submit that it is pure causalist dogma to base the denial on this. Someone taking this line does not keep an open mind about what might count as genuine perception: "It's very simple; when there's no causal link, I just don't call it perception!" No other consideration, such as the fact that the subject's description matches the world perfectly, will be able to persuade such a philosopher that we might, after all, have a case of perception. Indeed, this is the reaction of many to Dretske's example which was mentioned earlier, about someone who is able to describe all the details, including changes in position and orientation, of an object at the other side of a thick brick wall. However, it makes for uninteresting philosophy. If our *starting* point for the debate about the causal theory of perception is this dogma:

(CD) If there is no causal link between subject and object, then whatever may be going on cannot be perception

then, of course, nothing else but some kind of causal theory can come out as true. Wilkie puts the point thus:

> "Examples of veridical hallucination are [...] best seen as attempts (some better, some worse) to disguise the following assertion: if Jane cannot see John, then that is because John exerts no causal influence over Jane's visual impressions, experiences or sense-data. Stated baldly, it resembles a technical *definition* more than an argument." (Wilkie 1996 p.252)

CD forms the core of the causal claims about perception in this chapter – it is, as it were, the causal theory of perception stripped to its bare essentials, not making any claim about either the nature of the causal link between object and subject, nor about what the resulting effect state in (or: of) the subject has to be. But to refuse to budge from this dogma is in effect a refusal to do philosophy. If we are

to do any philosophy at all here, then we want a good reason to believe **CD** instead of adopting it unquestioningly.

One imaginable reason for believing in **CD**, to do with veridical hallucinations, might be that the absence of a causal link can explain the breakdown of veridicality at a certain point. Suppose for example that Fred describes the platypus on the other side of the wall correctly in all minute details, but fails to report its blinking an eye, or its walking to the left. He could *accidentally* have got his initial description right, but since there is no causal tracking, with every change in the object it becomes more unlikely that he keeps getting it right. Another example is that in which Fred's seeming to see the platypus is caused not by the platypus but by a hallucinogenic drug in his tea. Fred reports a platypus walking to and fro, and this *happens* to be a true description of what is happening. Again, the lack of causal tracking could be thought to explain why Fred's hallucination at some point ceases to be veridical.

Now all this may be true, but what exactly is explained by saying that it is a causal tracking that's missing? It suffices to say that these are cases in which certain counterfactuals aren't true. "If the platypus hadn't been there, Fred wouldn't have reported as seeing it" is false: in both examples, the animal can walk off without Fred reporting any change. But that counterfactual on its own suffices for distinguishing genuine perception from hallucination - if it's true, it is a case of perception, and if it's false it's hallucination. The falsity of the counterfactual explains why veridicality breaks down sooner or later as well as the absence of a causal link could explain it.

Still, one might object that this only holds causal theories at bay if the question as to how such counterfactuals are to be grounded is dodged. I will more thoroughly discuss the role of counterfactuals in the causal argument in chapter 5.

3.3.2 Unified defeating conditions and deviant causal chains

A peculiarity of the causal argument is that the claim that in cases of genuine perception there must be a causal link of a specified sort is arrived at by the observation that in *defective* cases such a causal link is *not* present, for example because it is prevented. But the mere fact that the success of some process can be causally prevented does not yet show that the process, if successful, must essentially be a causal process. For example: it can be causally prevented that two persons fall in love with each other - just keep them out of each other's way. Still, we would be reluctant to say that falling in love is essentially a causal process. The illusion corresponding to the causal theory of perception would be

the possible illusion that the two are in love with each other. The subjects may have correspondence by mail, and think they are in love, yet turn out to loathe each other when confronted. So this is not a real case of falling in love, we would say. Or take this example: an artist is making a painting, but because of bad lighting conditions his piece of art turns out to be no better than any child's amateurish mess. He thought he was making a piece of art, yet was causally prevented from doing so - but does this show that making a piece of art is of essence a causal process? The charge to the causal theorist wielding this type of argument is that merely by a judicious choice of examples he arrives at his causal claim.

An opponent to the causal theory may argue that in different cases of delusion or hallucination there is not one single unfulfilled condition, but a different one each time. In some cases there is an opaque object between the subject and the object he says he sees; in other cases, the object is not even there; in yet other cases, there is no light. Of course any concept of vision has to allow that such circumstances preclude vision; but it is not obvious that there is a unified defeating condition, namely the missing of a causal link between object and subject. However, suppose we use a heuristic principle to the effect that a satisfactory conceptual analysis would be able to unify such defeating conditions, instead of answering them on an ad-hoc basis. If we have a list of defeating conditions, then "(i)t is compelling to ask what unifies those conditions, what explains why the concept has just the conditions of application it does."[80] In other words, given the choice between a theory that does give a unified defeating condition and one that does not, it seems obvious that the latter is preferable. But now we need to consider two points. Firstly, does the causalist himself give such unified conditions? Secondly, is there any reason to think that a non-causal theory could not do so?

To start with the last question, it should be admitted that there are non-causal theories which don't bother to give unified conditions[81]. Once we recognise, however, that the non-radical anti-causalism with its list of conditions precluding perception is not the only possible alternative, there is no *a priori* reason why a non-causal account couldn't do better - or as well - in giving unified conditions.

As for the first question: is the causalist doing any better? His unified defeating condition for vision must be something like this:

[80] Child 1995, p.166

[81] see, e.g., Hyman 1992

(**UDC**) Only if an object has no causal role with regard to the subject's experience, what we have cannot be a case of vision.

In order to be unifying, **UDC** has to have the form of a necessary condition: that is why it starts with 'only if...'. But it is easily seen that in this form it simply is not true. It is well known that there are cases in which an object is causally affecting the subject, but that are not cases of vision. If in the dark you bump into a door, it is certainly causally affecting you, but you didn't see it. What we have here is the problem of deviant causal chains, which we already encountered in the previous chapter. Somehow these are not the 'right' causal roles.

The causalist will answer to this objection: "Of course I am aware that being causally affected is not a *sufficient* condition for vision. But what I was trying to give is a *necessary* condition, not a *sufficient* one." Most causal theories of visual perception are of form **NCA**, not **NSCA**, and there is a good reason for this. For if the causal theorist were to give a sufficient condition for perception he would have to solve the problem of deviant causal chains (on which more in the next section).

Bearing this in mind, however, we should come to the conclusion that talking about a *unified* defeating condition becomes rather gratuitous. To agree that conditions other than the mere presence of a causal link between object and subject need to be fulfilled in order for something to be a case of genuine visual perception, is by the same token to admit that there exist *other* sufficient conditions which defeat perception. Defeating conditions having to do with deviant causal chains do not fall within the causalist's 'unified' condition.

It may be thought that the additional defeating conditions could come to fall within the unified causal condition: once we have cracked the problem of deviant causal chains, we can formulate a genuinely causal unified defeating condition. Sceptical as I am about this possibility, I cannot exclude it. However, it should be noted that that is exactly the ambition of an **NSCA**-type theory, i.e. one that wants to give a condition for application of the concept 'vision' which is both necessary and sufficient. The argument here is not that such a theory cannot work, but rather that if that ambition is explicitly rejected then the argument from unified defeating conditions is not available either.

How much of a loss is this to the causal theorist? Let us consider again how **UDC** would be used. It is a piece of causalist ammunition with which to close the gap left between the **AFD**'s conclusion that there must be an extrinsic property promoting the mere visual experience to genuine perception and the claim that there must be a causal link of some kind between object and subject. There is a

good reason to think that the extrinsic property is a causal link rather than something else, since that alone gives us a unified defeating condition, the thought went. This line of argument has now been shown to be unavailable for **NCA**-type claims – as such, it is not a decisive objection against the causal ancestry claim, but a reminder that no decisive full argument has been given in favour of it. The other argument available for closing the argumentative gap for the causalist, as already mentioned when discussing **AFD**, is the argument from counterfactuals to which I will come back in chapter 5.

3.3.3 The nature of the causal link

Nothing has so far been said about the nature of the causal link that according to the proponents needs to be established between object and subject in order for the subject to perceive the object. From this it follows that the criticisms I have made so far are really directed against *any* theory which holds that, necessarily, there is some sort of link between object and subject, a link that serves to individuate visual experiences that are intrinsically the same. To object that any other sort of link than a causal one would be mystifying is therefore to misunderstand the issue.

In the following two subsections I shall consider objections to causal theories of perception which are specifically concerned with the nature of the causal link. First, I will discuss the problem of deviant causal chains. Then, I will discuss a problem that arises when a supervenience thesis is formulated.

3.3.3.1 Excluding deviant causal chains

As I have remarked before, most causal theories are **NCA**-type theories, and thus only aim to give a necessary condition for perception, not a full conceptual analysis of perception in the form of necessary-and-sufficient conditions. As a consequence, the problem of deviant causal chains is not a direct threat to causal theories of perception. However, in the previous section I have given reasons to think that the argument for the necessity of a causal link, on the basis of the **AFD**, implicitly also relies on its sufficiency. I think that the uneasiness about sufficient conditions motivates some attempts to define more narrowly what sort of causation has to be involved. Another source for the uneasiness may be the lingering thought that for a 'full' conceptual analysis non-causal conditions may have to be added, thereby diminishing the naturalistic appeal of the theory. For example, Grice, in his famous paper, does not rest content with stating mere necessary conditions, although he does not go further than making a case that we should be optimistic that it be possible to formulate sufficient conditions.

The problem of deviant causal chains is to exclude certain types of link which are of the 'wrong kind'; causal links the presence of which does not guarantee that we have a case of genuine perception. In other words, to confront the problem one has to specify which ways of causally affecting the subject are the 'right' ones. An object affecting the subject in the wrong way, such as the door that one bumps into in the dark, is just one problem. (It may be said that at least the door is perceived here, be it not by means of the right modality, namely vision. But if it is the concept of vision that we are after, how are we to narrow down? Condition (2), which states that the resulting state of affairs must be reportable as 'it looks to S as if...', at first glance, seems to do this. But it will not do, for it is circular: 'it looks to S as if such-and-so' already employs the concept of visual perception.) Another problem is that the object said to be seen is just one among many causes of our perception. E.g., when I look at some flower in the garden we do not say that I see the sun, although the light emitted by the sun is one of the causes of my perception. Similarly, we do not say of a bat that it perceives itself, although its perceptions are at least partially caused by the sonar-waves that it emits.

One attempt at narrowing down the valid causal links was already done in the **SDC** as I formulated it, namely in the condition that the object must cause in the subject a state which can be reported as "it looks to S as if an O is there". But that condition cannot do any work, for it circularly refers to 'looks'. Circularity is also a threat to other proposals. Grice, for example, proposes:

> "..for an object to be perceived by X, it is sufficient that it should be causally involved in the generation of some sense-impression by X in the kind of way in which, for example, when I look at my hand in a good light, my hand is causally responsible for its looking to me as if there were a hand before me, or in which ...(and so on), *whatever that kind of way may be.*"[82]

Grice, then, defines perceptual causation by means of paradigmatic cases, leaving open a blank to be filled in by specialists. But as Dancy[83] remarks, there is a danger here. What is the count as the same "kind of way...whatever that way may be"? Not all circumstances will be the same in different cases, so we have to decide which ways are relevantly similar. This work cannot be left to the neurophysiologist :

> "If they found an apparently perceptual belief that was caused in a completely new way, who would decide whether the way was relevantly similar to previous more well-trodden ways? Neurophysiologists have no special right to make this decision. And we would decide it, not by considering the degree of similarity between different causal histories, but

[82] Grice 1961 p.71, my emphasis
[83] Dancy 1985 pp.174-175

> by considering directly whether we wanted to treat this belief as perceptual."
> (Dancy, 1985 p.174)

The disagreement between Grice and Dancy in the end boils down to a disagreement about whether mental kinds such as perception are natural kinds. If something is of a natural kind, then it is an objective fact, i.e. independent of our decisions, whether a specimen belongs to it. We recognise a natural kind initially by some paradigmatic features, such as e.g. colour and specific gravity in the case of gold: all other details and possibly hidden structure are then for science to be discovered. So if perception were a natural kind, then we could indeed learn more about the kind of causal link involved by just investigating the paradigmatic examples in depth. This discussion lies at the heart of the causalist - anti-causalist opposition, and I will delve deeper into it in chapter 5.

3.3.3.2 Causation and supervenience

Brian McLaughlin[84] argues that any causal analysis of perception implies the following **f**alse **i**mplied **s**upervenience **t**hesis:

(FIST) If an object and a person's experience bear the same causal connection to each other (in relevant aspects) as do a second object and a second person's experience, then the first person perceives the first object if and only if the second person perceives the second object.

FIST is motivated by what McLaughlin thinks are common assumptions of causal theorists:

> a) Perceptual causation supervenes on at least one of the kinds of micro-causal process by which a perceived object can cause a perceptual experience.
>
> b) A micro-causal process by which an object causes a person's experience can be an instance of perceptual causation solely in virtue of its intrinsic properties.

So whatever kind of micro-causal process is going on in a specific case, it causes the perception in virtue of its intrinsic properties; and if in a similar case exactly the same micro-causal process is going on, then the perception must also be the same.

However plausible this may sound, according to McLaughlin it is not true in certain kinds of cases. Consider cases in which perceiving part of an object counts as perceiving the object itself. This will only be so under a certain set of

[84] McLaughlin 1984

conditions; these, however, do not supervene on the micro-causal processes by which the perception takes place. What we have here are simply *conventions* governing proper attachment that can change without a change in the micro-causal process. For example, seeing one brick of a building would not count as seeing that building; but seeing a small part of the Atlantic Ocean does normally count as seeing the ocean. Such conventions can easily be imagined to be different. Another type of case is that in which seeing something caused by an object counts as seeing that object: e.g., we say we see a car approaching when all we see is moving lights through the dark.

McLaughlin's arguments show that perceptual causation cannot supervene on micro-causation. His next (implicit) step seems to be that therefore 'perceptual causation' is misleading nomenclature - what is involved in perception is not a causal process. However, couldn't a causal theorist of perception object that micro-causation is to small a supervenience base for perceptual causation, and thus that McLaughlin's conclusion is too radical? To do so, an alternative supervenience base needs to be proposed for perceptual causation. This base would apparently have to include facts about the common practice of perceivers: what matters are the rules and conventions that are followed. Specifying such a supervenience base would amount to naturalising the rule-following practice of the relevant perceivers. Stopping short of a discussion of the thorny topic of rule-following, I want to suggest that this is a daunting prospect.

Another causalist reply to McLaughlin could be that what we *say* to be seeing is not necessarily what we *are* seeing. Even if we say that we see the Atlantic Ocean, that does not make it the case that the whole Atlantic Ocean is the object of our perception. Why should our linguistic practices tell us anything about the objects of perception? Why would there be a problem in saying that we see the ocean while, in a strict sense, seeing only part of the ocean? However, to put the issue in these terms is misleading: for what we were after was an understanding of the concept of perception. Therefore, we must have good reasons if we want to say that what we ordinarily refer to as the objects of our perception diverges from what they really are. To insist that linguistic practices must be irrelevant because otherwise supervenience of perceptual causation on micro-causation could not be true is to be biased in favour of a causal theory.

3.3.4 Reliability

We now come to the third argument that I mentioned at the outset of our discussion of causal theories of perception. It is of Kantian origin, and can be found in Strawson:

> "For we think of perception as a way, indeed the basic way, of informing ourselves about the world of independently existing things; we assume, that is to say, the general reliability of our perceptual experiences; and that assumption is the same as the assumption of a general causal dependence of our perceptual experiences on the independently existing things we take them to be of." (Strawson 1979, repr in Dancy 1988 p.103)

I want to concentrate on the inference from the general reliability of our perceptual experiences to a causal dependence of our perceptual experiences on the world. It certainly seems reasonable to assume the general reliability of our perceptual experiences, but does causal dependence follow from this? The crucial notion here is 'reliability', which is, I take it, a normative notion. Can we, on the basis of this normative feature of perception, draw conclusions about the metaphysics of experience? I argue that we cannot.

Imagine the following case: a friend of yours claims to be clairvoyant. Sceptical about his predictive powers, you ask him to tell you what will happen to you tomorrow. He tells you in great detail that you will receive a letter from a certain review that they finally decided to publish the article you sent them; that around lunchtime you will encounter a friend that you haven't seen for two years; and so on. Although you are in a sceptical mood next day as you climb out of bed, everything happens exactly as predicted. Thinking that this must be a cosmic coincidence, you decide to consult him again. To test things out, you even try this time to prevent the predictions from coming true - but to no avail. Now if this were to continue for a substantial period of time, would you not be forced to conclude that his predictions were reliable, even if you hadn't got a clue how he went about obtaining them? What I want to argue is that one doesn't need to know **how** something can be reliable in order to be able to conclude that it **is** reliable. Suppose you found out that your friend was very rich and influential, and thus able to make his predictions come true: that would certainly explain how the predictions could be reliable. "Of course his predictions are reliable!", you exclaim, trying to suppress your feeling of having been cheated. Nonetheless, you won't be able to complain to your friend that his predictions were not reliable; for they were.

You don't need to be a car mechanic, nor is it necessary that you have ever looked under your car's bonnet, in order to be justified in concluding that your car is reliable, simply on the basis of its performance until today. And

conversely, you may invest billions of pounds in designing a new aircraft, and use only the best materials and technicians to assemble it, and still it may crash on its first test flight, or explode in the air. A new model aircraft or motorcar may be marketed as having an excellent chance of being reliable, yet nothing is really known about its reliability until it has been in use for some time.

Coming back to the case of perception: we have good reason to assume that our perception is generally reliable in informing us about the world, given our success in navigating through it and acting on it. But making this assumption does not require us to know how this can be so, or why it is so, just as you don't need to look under the bonnet of your car to say something about its reliability. Now just as knowing that your car is well-designed and properly maintained may give you good reason to believe it to be reliable, the knowledge that your perceptions are causally dependent on the world may *give you reason to believe* that they are reliable. But this observation does not help us to the conclusion that there must be a causal link if there is to be reliability. For inferring causal dependence from reliability, as Strawson does, requires that your perceptions be reliable **only if** they are causally dependent on the world. After all, claim **NCA** is in search of a necessary condition, not a sufficient one.

We want to know, then, whether the reliability of perception could be due to something else than causal dependency of visual experiences upon their objects. One specific form of this question is: is occasionalism a conceptual possibility? This question is discussed by Child[85]. Occasionalism is the view that God (or whoever) simultaneously produces both things in the world and experiences in us which truly tell us about them. Child challenges the occasionalist to tell us "What reason...there [could] be for saying that experiences...really have contents concerning things in the world?". His position is that we already have to know what relation holds between material objects and experiences: otherwise, we have no reason to assume our experiences to be reliable. The argument is then completed by saying that, given occasionalism, we have no way of knowing that it is true. I think, however, that this argument is wrong in two ways. Firstly, in the last step a causal theory of knowledge seems to be assumed; but it seems plausible that a theory of knowledge would be based on a theory of perception, and therefore this reliance must be a form of question-begging. Secondly, and more importantly, I argued earlier that we don't need to know how it can be that our perceptions are reliable in order to be justified in assuming that they are. Knowing, as we do, that our perception helps us in getting around is as good evidence - or maybe even better - than knowledge about the mechanism of

[85] Child 1994 p.170 ff

perception, just as one is justified in asserting that a car is reliable if it has never let you down. And perhaps even the emphasis on evidence here is not quite right. It is one thing to hold that in order for something to be perception it needs to be (generally) reliable (in 'telling us what the world is like'), and quite a different thing to say *that we must have a justification for assuming that* it is reliable. In order to be a good car, it needs to be reliable – that's just one of the things that makes a car into a good one – but there's no demand that we must be justified in assuming it to be reliable. Compare this to: in order to be a minor, one needs to be less than 18 years of age. It is not necessary for somebody to be a minor that we be justified in assuming him to be less than 18 years of age. Similarly, the first claim seems to do the conceptual work in the case of perception.

Given what we know about neurophysiological processes are involved in perception, it seems unlikely that occasionalism be true. (But it is not impossible, since we do not know what the relation is between our brain states and our experiences.) However, there seems to be no conceptual reason why it could not be true, and there still be perception in this world. In any case, even if occasionalism were to exclude perception, it would not be because of lack of reliability. If God were to cause our experiences, surely we could rely on Him doing that such that our experiences tell us what the world is really like?

Name of position (proponent)	Classification as causal theory of mind	**Objections ->**	Empirical contingency, not conceptual necessity	Indistinguishability and the metaphysics of sensa	Explanatory of possibility of 'perfect hallucination'?	Reliability not implying causal relation	Unified defeating conditions	Deviant causal chains; role of the expert
Sense-data causalism (Paul Grice)	N(S)CA		#	*	#	-	#	*
Disjunctive causalism (William Child)	FPCE / NCA?		?	*?	#	#	#	-
Reliabilism (P.F. Strawson)	NCA / FPCE		?	?	?	#	?	-
Adverbial theory (Michael Tye)	NCA		?	-	#	-	?	-
Experientialism (Alan Millar)	NCA		?	*	#	-	#	-
Disposition-to-believe (Smith & Jones)	NCA		?	*	#	-	?	-

* constitutes an objection to the position;
\- does not affect position, or is satisfactorily answered
\# undermines an argument in favour, or motivation, of the position
? position not sufficiently worked out to assess impact of criticism, or controversial

3.4 Summary and conclusions

In this chapter I have looked at causal claims about the concept of perception. The most important type of claim is one about causal ancestry (**NCA**), and the most prominent argument in favour of that claim the argument from delusion. It has turned out to be a very problematic argument. Firstly, it simply does not establish a causal conclusion. Secondly, it is based on an epistemological position (the common factor approach or 'conjunctivism' that leads to scepticism. The disjunctive theory of experience is not obviously successful in getting rid of the epistemological trouble, and although compatible with **NCA**, robs the argument from delusion of any argumentative teeth it had left.

Most of the literature on the causal theory of perception was found to be concerned with dealing with the epistemological predicament into which the argument from delusion has led us, without attempting to strengthen the causalist's case. Perhaps William Child is an exception to this. He distances himself from the causal ancestry claim and makes an explanatory claim instead, and explicitly offers two more arguments. The argument from unified defeating conditions is problematic because of deviant causal chains, even though it supposedly disavows the sufficiency claim **NSCA**. And the argument from reliability, which departs from the idea that perception must be reliable in telling us about the world, was found wanting because reliability does not imply a causal link.

The causal ancestry claim about perception, it can be seen from the above schema, does not stand as strongly as is commonly supposed. Admittedly there are some more resources for the causalist in considerations about counterfactuals, and about natural kinds. Discussion of those arguments takes place in chapter 5. But what can we learn from the discussion of causal claims about action when we juxtapose it to action? That is the subject of the next chapter.

4 Cross-fertilisation and disanalogies

In the previous two chapters I have discussed how well the arguments in favour of various causal claims with regards to action and perception have stood up, and which problems these claims have to confront. Now I want to consider what the main differences and similarities are in the discussion over the various causal claims with regards to action and perception. The purpose of this chapter is twofold: firstly, to try and identify what the main recurring themes are in the discussion over causality and mind. Secondly, to see how much cross-fertilisation there can be between the discussions over the various mental concepts.

4.1 Varieties of causal claim

In the previous chapters we have seen that various different claims with a causal character can be made regarding the mental. The objections which I formulated applied sometimes to one such claim, then to another, and sometimes to several. In the attempt to assess the extent to which there can be cross-fertilisation between the discussions in previous chapters, we need to see whether the same type of claim is being made for the different mental concepts.

In the philosophy of action, it could perhaps be said that the dominant causal claim is one about explanation. This is the claim that folk-psychological explanation of action in terms of reasons must be causal explanation because otherwise we could not account for its strength, and which I gave the acronym **FPCE**. As we have seen it is not so clear that this claim can stand on itself. Could we still hold that causal explanation played such a central role in folk-psychology if we denied that mental states are causal states (**FPCS**), in the sense that mental states are token-identical with brain-states located on the nodes of the causal network? Can philosophers who deny such token-identity give an account of how the causal explanation is made true, while holding on to the claim that the explanation is distinctively causal?

While not needed to support the claim itself it is attractive to expand the explanatory claim into a claim about what the conditions are for something to be an action. If it is claimed, firstly, that actions are those things that lend themselves to causal explanations in terms of reasons, and secondly that causal explanations are made true by explanans and explanans and explanandum referring to causally linked events, then we have ended up with the claim that a necessary condition for application of the concept of action has to do with its

causal ancestry (**NCA**). This, I argued, is Donald Davidson's position. Volitionists make a claim about application of the concept (**NCA** or **NCP**) without making this detour.

In the philosophy of perception, things are different. There the main claim is of the type **NCA**. This might be explained by the sceptical worry that we might be hallucinating when we think that we are perceiving; therefore the need is felt for a criterion by which to distinguish one from the other. Although Grice cautiously ventures in that direction, only few think that it will be possible to formulate a sufficient as well as a necessary condition for application of the concept of perception in terms of causal ancestry. There seems, then, to be a clear difference between perception and action in the type of causal claim generally made, even if there is also a clear overlap.

A few authors of causal bent try to strengthen their position by pointing out parallels between action and perception, and it is interesting to take a see how they deal with this apparent difference in slant between causal claims about the two concepts. Hornsby, who proposes an **NCP**-type claim about action does not have this problem. She writes:

> "To describe an event as a perception (a perceiving of something) is to describe it in terms of its causes: to describe an event as an action is to describe it in terms of its effects. ... [I]t has been necessary to impute a number of confusions to other philosophers' conceptions of action; but if I can show that these correspond one to one with as many potential confusions about perception, and if it is clearer that these are confusions, then this may lend further support to the account of action." (Hornsby 1980, p.111)

She explicitly identifies Grice's account of perception as the counterpart 'like a left-hand and right-hand glove' of her own account of action. If she is right about this, then we should also expect both accounts to be affected by the same objections.

Child proposes an **FPCE**-claim for action while explicitly rejecting the **NCA**-claim; as a consequence he must also defend an **FPCE**-claim about perception. He formulates it cautiously: the object of perception is supposedly *causally responsible* for its looking thus and so to the subject[86]. What he wants to achieve, in order to formulate a causal theory of perception that does not fall prey to scepticism, is that the object itself is essentially part of the visual experience. His thought is that the object cannot both cause and at the same time be part of the experience, but that it can causally explain the experience while being part of it[87]. However, can we make sense of the object (causally) explaining the visual

[86] Child 1994 p. 141
[87] Child 1994 p.160-162

experience? Or rather, if we try to draw the parallel more fully: does an explanatory (folk-)practice exist in which what people seem to see is explained in terms of the objects of their perceptions? And if so: is there an issue over what accounts for the strength of such explanations?

When I perform some action, it may be perfectly clear to an observer why I do what I do. Suppose, for example, that the doorbell rings and I go to open the door; what I do will not normally be surprising to you. But there may be contexts in which it is, for example when I have just been telling you that I expect somebody to come whom I do not want to see. And we all know how small children, at some stage in their development, will not tire of asking about everybody why they do this or that – usually trivial actions. So in many cases (but if the point about arational actions was right, not all) some explanation may sensibly be asked for. But now suppose that I seem to see a platypus. Does it make sense for you to ask me *why* I seem to see a platypus? And supposing that there is a platypus in front of me: does my pointing this out to you answer your demand for an explanation? If you asked me the question in the first place, it must have been because you thought that the obvious public fact of the platypus' presence did not explain it. When you ask for an explanation of my opening the door, even if you heard the doorbell as well as I did, I can, in order to explain my action to you, place the doorbell's ringing against a background of reasons that might not have occurred to you. But the envisaged explanation of my seeming to see something cannot be like that, for the simple reason that it is not a reason-explanation: people don't have reasons for having visual experiences – they just have them. Why should the envisaged explanation be a reason-explanation? It is likely that we can give some causal explanation in terms of physics and neurophysiology. But, firstly, it is doubtful that that provides you with the kind of answer you wanted, and secondly, such a scientific explanation is constructed *as* a causal explanation. That is to say, supposing that it is an acceptable scientific explanation, there is no room for a question analogous to that about reason-explanation, namely: does the force of this explanation derive from its being causal? In short: I do not see how an **FPCE**-claim about perception analogous to that about action can be set up meaningfully. Here we have a clear disanalogy between action and perception.

For Child all this means that he cannot support his causalism about action unless he bites the bullet and admit an **FPCS** or **NCA** claim about action. In general it means that we have to treat analogies between mental concepts with caution when discussing causal claims, and verify that the claims are of the same type. It does however not invalidate the strategy of drawing the parallels, because the

different causal claims are more often than not related. But it does make matters more complicated.

4.2 Objections to causal claims

4.2.1 Deviant causal chains

If we want to make a claim of type **NCA**, i.e. if we want to individuate instances of it from instances of other concepts by saying that they have a characteristic type of causal ancestry, then we will have to say something about what causal ancestry that is. Since, presumably, anything is caused by something or other in one way or another, it needs to be specified *by what* exactly those instances that we want to distinguish from others are caused, and *in what way* precisely. The causal theorist told us that actions are those movements caused by primary reasons (desire + belief), and cases of perception are those in which the visual experience is caused by the object of it. I will discuss the nature of the cause in the next subsection. Here I want to focus on the second bit that the causalist needs to fill in. There are many well-known examples in the literature which show that simply having the right cause does not suffice: recall Chisholm's nephew, Davidson's climber, and the gardener who does not see the sun by looking at the flower.

There are two possible ways of reacting to these examples. One is to argue that they do not pose a problem, since the project was to find a *necessary* condition for something to be a case of action or perception: that that condition is not *sufficient* does not matter. As I have pointed out in section 3.3.2, this approach undercuts an important argument for causal theories. The argument from unified defeating conditions said that a reason for thinking the claim about causal ancestry to be central to the concept is that the absence of causal ancestry unifies the defeating conditions for instances being of that concept. It is undercut because it now becomes clear that not all defeating conditions are 'unified' - because the ones having to do with deviant causal chains are left out. On the 'it's only a necessary condition'- approach it is simply admitted that the causal theory does not have the resources to sift deviant causal chains out as defeating conditions: but in that case we can at most speak of partial unification of defeating conditions, and the conviction carried by the above argument for causal theories evaporates. If the causal condition cannot achieve complete unification, what reason is there to believe that it is relevant to such unification at all?

The other approach is to try to spell out conditions which have to hold if the causal link is to be of the 'right' sort. This is the approach mostly favoured in the philosophy of action. This kind of proposal, it seems, comes in three sorts. The first sort of proposal simply does not work: these are the various extra conditions to which counterexamples can be found. The second sort of proposal is open-ended: it delegates to (future) neurophysiology the task of finding out what the type of causal link involved is (whether in action, or perception). But can empirical research provide the answer to the question of what is to fall under a concept? That seems doubtful: on what grounds can the neurophysiologist, when confronted with a certain type of causal link, decide whether what he has under consideration is (e.g.) a case of perception - especially if it seems a borderline case? Rather, the concept would need to be clarified first: the scientist needs to be told what to look for. The only case in which this strategy would work is if mental concepts were natural kind concepts; I will argue in chapter 5 that they are not. The third sort (e.g. Bishop's "sensitivity strategy", see p.43) formulates certain counterfactual conditions on the relationship between cause and effect. However, such conditions could also be true of a non-causal relation between events. I argued that since they alone suffice as criterion for the distinguishing work the causal condition was thought to do, it becomes unclear why we should need the causal condition as opposed to a mere counterfactual condition.

Opinions differ on how serious the problem of deviant causal chains is. Many think that it is serious, but not hopeless: it is just more work and creativity that is needed. Davidson, though, is pessimistic: but he does not consider the problem to be central to his philosophy of action, probably because his first concern is not with the **NCA** claim but with the explanatory (**FPCE**) claim. Others try to show that deviant causal chains are a problem which in principle will resist resolution. Wilson[88] shows that any strengthening of the conditions accompanying the causal link is open to a yet more refined counterexample; and I have suggested (p.40) that if we make the link between intention and action too immediate we get into trouble with the logical connection argument instead. Moya[89] suggests that intentional action has an ineliminable normative component, which is not capturable in terms of causation:

> "That the relation between reasons, intentions and intentional actions is normative and that normativity itself has to be given genuine efficacy in prompting our intentional actions is strongly suggested by the fact that the opposite assumption cannot yield a correct and complete analysis of intentional action, as the problem of wayward causal chains seems to show. If we dig deep enough into this problem we are naturally led to the

[88] Wilson 1989
[89] Moya 1990

> conclusion that a general, scientific concept of cause is not able to capture the structural relations involved in human intentional action..."(Moya 1990, p. 168)

The problem of deviant causal chains seems to be a problem for any attempt at defining a concept by its causal ancestry: examples to try are the concepts of music, bullying, education, politics. Therefore it is a general problem for this kind of approach, although there is a possible exception for natural kind concepts; I discuss this point in chapter 5. Where it concerns mental concepts, the point should not at all be new to philosophers, for the discussion about Gettier-cases in the theory of knowledge reflects much the same point[90]. Demanding a causal link has seemed to many philosophers the obvious way to deal with objections to the tripartite analysis of knowledge; but that 'obvious solution' has given rise to a controversy on its own.

4.2.2 The cause

A causalist who holds a claim of type **NCA** about action has to conceive of willings, tryings or primary reasons as the kind of thing that can be a cause. However, on the assumption that these mental states are similar to beliefs, there is a potential problem indicated by Moore's paradox. I discuss this in the first section below. Then I go on to discuss a problem for both **NCA** and **FPCE**-claims, namely that an appropriate causal-explanatory item need not always be present.

4.2.2.1 Moore's paradox

Moore's paradox starts from the thought that one can have false beliefs: in other words, believing that P, and P being the case, are independent. One can say: "He believes that it is raining - but it isn't raining", without any contradiction. But the case is different where it concerns oneself: one cannot sincerely say, "I believe that it is raining", and add in the same breath, "but it isn't raining." To say so would be inconsistent with how we normally use the term 'belief'. So there is an asymmetry between the 1st and 3rd person case.

If we are conceiving of beliefs as states-in-the-head this asymmetry is puzzling. For if we do so, then it seems that we must be talking about what the insides our heads are like when we report what we believe. And if that were the case, why should it make any difference whether the world is indeed as we believe it to be

[90] Gettier 1963

or not? Why should we *not* report that we believe that p, and assert in the same breath that not-p? There seems to be a definite clash between our ways of talking about belief and the assumption that beliefs are inner causal states. Jane Heal[91] argues that it becomes impossible to account for the paradoxical nature of the Moorean thought if one accepts the following two claims:

(1) the linguistic behaviour of the word 'believe' is to be explained by unpacking the truth conditions of claims about beliefs, and

(2) these truth conditions are filled in on the assumption that belief is something which goes on within a person but is independent of how things are outside him or her.

It seems plausible to think of Moore's paradox as providing us with a condition that must be satisfied by any account which aims to analyse or clarify the concept of belief. The account subscribing to the above two claims – Heal calls it functionalism, and my **NCA**-causalist will subscribe to it, too, unless he argues that the shape of the account of the concept of belief is very different from other mental concepts – fails to explain the *conceptual oddity* of the Moorean thought. At most, it allows that somebody's having a Moorean thought is an unusual situation; somebody having such a thought has, on the functionalist account, a strange psychological make-up, but the oddity does not arise from the nature of the concept itself.

According to Heal, there are general reasons for thinking that any functionalist approach must run into the same trouble on this point, and moreover that this is a serious problem for functionalism[92]. It does not do justice to the lesson which we should draw from Moore's paradox, namely that "the real shape of the concept [of belief] is one in which criterionless first person ascription and behaviourally based third person ascription are inseparably linked"[93].

In the current context however we need to realise that the point made here does not establish that beliefs cannot be causal items at all: the subject matter of the concept 'belief' may well consist of causal states. In other words, Moore's paradox gives us no reason to deny claim **CE** (see chapter 1, and for a further discussion 5) about beliefs. However, it does follow that no full conceptual analysis of belief can be given in terms of the causal relations of belief-states. This means that a functionalist account of belief has to be supplemented, or replaced by another account.

Does a similar 'paradox' arise for other mental concepts? Let us consider perception. Can I say, "I see a cat, but no cat is there"? That certainly seems to

[91] Heal 1994

[92] Heal 1994 p.17; a similar point is made by Collins 1987.

[93] Heal 1994 p.23

be paradoxical; one would at least be tempted to retort, "how can you say those two things at the same time", or, "do you not mean that you *thought* that you saw a cat?" There is however, unlike in the belief-case, no asymmetry here. It makes no difference to this exchange whether it is about what *I* perceive, or about what *somebody else* perceives. Other than with belief, seeing is a concept that does not allow for a discrepancy between what someone sees, and what the world is like. With visual experiences, where there can be such a discrepancy, the paradox does not arise because in saying that one has a visual experience one does not make any epistemic commitment as to what the world is like; and therefore, again there is no asymmetry between the 1st and 3rd person cases. Moore's paradox, we may conclude, seems to be specific to belief. We do however need to bear in mind that, since belief plays such a central role, causal claims about other concepts may assume that beliefs are causal items. The causal ancestry claim about action is a case in point.

4.2.2.2 Appropriate causes

In discussing arational actions in section 2.2.6, we saw that the kind of cause to be postulated for intentional action may not in all cases be present, namely in those cases where actions are not done for any reason or not with any intention, not even under another description. If such cases are indeed, as I argued, to be considered as instances of action then a Davidsonian loses the unification in the kind of cause needed to distinguish actions from mere events. Is there an analogous problem for the causal ancestry claim about perception?

The discussion in the previous chapter was - in accordance with tradition - almost exclusively about the perception of material objects. These are, however, not the only things we visually perceive. We do say things like: "I perceived a very tense atmosphere in the room", "I saw that the bacteria had multiplied rapidly", "I saw people streaming out of the stadium", "I saw him struggle to open the pot of pickles". These are all cases in which what is perceived is not simply a material object, although at the same time they are very unmysterious in that they do involve material objects. If we want to make sense of seeing anything other than objects, then the causal story starts to look very simple-minded. If a causal ancestry claim is to be fleshed out for all visual perception, what might the causing object be in the examples above?

The causalist could reply that, properly speaking, it is events which are causally linked. When it was said in discussions about perception that "the cup causes my visual experience", that was really short for "The cup's standing there causes my

having a visual experience". In the cases above we can still identify events that might act as causes. This reply certainly fits well with Davidson's anomalous monism and his account of causation. But this move is not much of a reply: more would need to be said about the kind of event that is admissible, because for an ancestry claim not any event will do as a cause. The problem in the case of actions is to find a cause both of the specified type - a primary reason - and with contents somehow congruous to the action. Similarly, with the perception of things other than material objects, the problem is to find a causing event that can be given a description related to the content of the perception.

Another causalist reply might be that the alleged counterexamples are in fact based on inference. For example, we don't perceive a tense situation, but only a frown on this face, a raised or lowered voice there, etc., and from that we make an inference. However, are such things 'simple' or do they need to be inferred from, e.g., perception of present eyebrow position together with knowledge of usual eyebrow position? Pursuing this reply to its logical consequences leads to an absurdly restrictive view of what can be seen; I discuss this position in section 4.2.5.2 below under the name 'epistemic atomism'.

4.2.3 Mental causation and supervenience

According to **NCA**, an action distinguishes itself from a mere movement in virtue of the fact that it has been caused 'in the right way' by a specific mental event (the primary reason). A visual experience is genuinely perceptual rather than hallucinatory in virtue of the fact that it has been caused 'in the right way' by the object that it is about. The insistence that mental concepts such as action and perception essentially involve a causal link as discussed appears to be the main motivation for holding that there must be mental causation.

Mental causation is, nowadays, hardly ever discussed without the obligatory prefix of 'the problem of'. It leads to a problem because it is hard to see how the mental can have any causal autonomy when it is determined by (supervenes on) the causally closed physical, and massive overdetermination by mental and physical causes is rejected. Combining a dualism of properties with a monism of substance appears to render one set of properties causally irrelevant or epiphenomenal, and since from the naturalist perspective we can hardly give up physical causes it must be the mental causes that lose out. I discussed all this at greater length in section 2.2.5.

How the problem of mental causation affects the causal theory of perception is less appreciated. In section 3.3.3.2 I discussed McLaughlin's thesis that

perceptual causation cannot supervene on macro-causation. It cannot do so, because what we see (the content of our experience) and the micro-physical cause of our experience (or rather the micro-physical state identical with it) are not always straightforwardly identical. One way of putting it is this: the object of perception is reflected in the content of the perceptual experience. However, this content is context-dependent, whereas it is hard to see how micro-physical causation could be. Clearly, again, fitting the content into the causal story - accounting for the congruity between physical causation and content-causation - is a problem.

Yet the causal ancestry theorist about action and perception needs mental causation. Actions are actions because they are caused by the primary reason for them, not because they are caused by some neurophysiological state which happens to be identical with the primary reason. Similarly, a visual experience is genuinely perceptual because it was caused by the object that it is an experience of, and not because it happens to be token-identical with a neurophysiological state which was caused by the object that the experience is of. For this reason a problem about content is going to arise with any mental concept that we try to define by means of an extrinsic causal link, while holding on to the naturalist outlook. Hence the increasingly desperate attempts to somehow find a place for content in the natural world.

4.2.4 Regress

The causal ancestry claim about action is based on the assumption that we have an item the intrinsic properties of which do not differentiate it as an intentional action or mere bodily movement; it is, in itself, neutral. That consideration led to the thought that the difference must lie in a movement's causal ancestry. But what kind of cause is required? If we demand that the cause itself be an action, then we just duplicate the question, starting a regress: for what makes *that* cause an action rather than a mere movement or happening? This is not to say that actions cannot be caused by actions, only that it still leaves the question: what made the first action in the chain (the causally basic action) into an action? Tracing back the causal chain leading up to the action that we are interested in, it can come to a stop in either of two ways. Either the first element is some sort of 'pure' action, not supervening on - and therefore not implying the occurrence of - a mere event or happening. This is the traditional volitionist solution, which has two problems: is the 'pure action' an entity that can be coherently fleshed out in a materialist ontology? And in virtue of what is the volition an action - is it the exception to the rule that what makes something into an action is its causal

ancestry? The modern volitionist theories which I discussed come up with clever answers to these questions, but at a price - they cannot, in the end, offer unproblematic and convincing solutions. Or - this is the second way to stop the regress - at the beginning of the causal chain we find an event which is not itself an action. But not any old kind of event will do; for mere bodily movements are also caused by some sort of event. Obviously, if we want to distinguish intentional actions from mere bodily movements by their causal ancestry, the sorts of events that do the causing must be different. According to Davidson's anomalous monism, the distinctive character of the cause of an intentional action lies in its being the primary reason for that action. This is not unproblematic either: mere bodily movements can equally be caused by primary reasons, so the cause being of the right kind does not give any guarantee. Another worry is whether primary reasons are the right sort of thing to be causes. The basic worry for causal-ancestry theories could be formulated in this way: what reason is there to think that an event's cause can determine what sort of event the effect is?

Turning our attention now to the causal theory of perception, there is a worry of regress there, too. However, it is not a 'source'-problem as is the case with action: the beginning of the causal chain is rather well-defined, starting with the object of perception affecting the subject. Thus, in the case of visual perception, the object in question will reflect light falling on it; this light will enter the eye, go through the lens and fall onto the retina. The receptors on the retina will then stimulate the optical nerve, and the optical nerve will eventually cause some event in the visual cortex, this in turn will cause various other events. Where in this causal chain do we find the perceptual experience? That is, at which point does the chain result in the item that we were interested in? Where in the case of action it is unclear with which event the relevant causal chain starts, in the case of perception it is unclear what terminates it.

It will not do to say that the visual experience is that event which the subject is aware of; for which event is that? The awareness here would not be a perceptual awareness, because - as is well known from discussions about sense data - such an 'inner spectator theory' would lead to an infinite regress. With action, we somehow had to find a first element in the causal chain that could be proclaimed an action without begging any questions; and now we see that with perception, finding an end to the causal chain without begging any questions is a tall order. The 'inner spectator theory' seems to opt for a similar 'solution' for terminating the causal chain as the volitionist did for actions: to postulate a 'pure' action (volition) or a 'pure' perception (visual experience). The analogue of the other option for terminating it would be to postulate of some brain event or other that it

is the visual experience, presumably on the grounds that it has certain effects (e.g. brain states identical with perceptual beliefs.) But looking at the causal chain of brain events will not straightforwardly enable us to decide which events are the visual experience, and the perceptual beliefs.

4.2.5 Mota and sensa

4.2.5.1 'Indistinguishable'

The causal ancestry theorist aims to offer an account of the distinction between mere happenings or mere bodily movements and intentional actions; or between genuine perceptions and hallucinatory or delusory experiences. According to this account, there is no intrinsic difference, so we need to look at extrinsic features of these states or events. The distinction that we are after has to do with how the phenomena in question were caused.

This reasoning is built upon the thought that there are neutral items which do not have any intrinsic properties differentiating them as instances of one or the other: visual experiences are neutral between delusory and perceptual experiences, bodily movements neutral as between mere movements and intentional actions. Now recall that two things transpired from the discussion on indistinguishability in the previous chapter. Firstly, if two things are indistinguishable they must be items of awareness, and thus we would commit ourselves to an 'inner spectator' theory. But this leads to an infinite regress, and is not giving an account of perception at all. Secondly, there is a motivational point. Even if one resisted – as the disjunctive theorist of visual experience does - that in order to have indistinguishability one does not strictly *need* neutral items, one may be challenged to explain how it can be that hallucination and genuine perception can be mistaken for one another. However, I have shown that such a demand had better be rejected. When two things are indistinguishable we have to say in what respect they are so (though it is logically possible that they are indistinguishable in *all* respects). I argued (p. 87) that when we say that cases of perception and delusion are indistinguishable to the subject, they are so in respect of their involving the same visual experience. But that they involve the same visual experience does not in any way explain why genuine perception and delusion are indistinguishable, or how the subject can mistake one for the other. That is just a fact about perception and delusion which is not open to further explanation. If there were not this possibility of mistake, then it simply would not be a case of delusion – therefore it would be odd to still demand an explanation of it.

Are the neutral, indistinguishable, elements in the causal theory of action as problematic as sensa? Let us, without prejudging their objectionability, call such neutral bodily movements *mota* for short. One might think that there are at least the following two disanalogies between sensa and mota.

Firstly, one might argue that mota are physical events whereas sensa are mental events. Mota are just bare movements that can be described using the language of physics. As such they are neutral - physics cannot make the difference between mere movements and intentional actions - and we have no problem in picking them out. How can we refuse items in our ontology that so obviously exist? However, to reason in this way would be a mistake: there is a confusion here between concept and subject matter. I am not denying that both intentional actions and involuntary movements (e.g.) can both be identical with (physical) movements. But insofar as within the vocabulary of physics we cannot distinguish between mere movements and intentional actions, we cannot talk of *neutral* bodily movements either. Or rather: in the vocabulary of physics *all* movements are neutral, and therefore within this vocabulary there is no room for a concept of items that are neutral between two other categories that cannot even be distinguished. Neutral bodily movements (mota), as well as visual experiences (sensa), the conclusion should be, are mental items. More precisely, 'motum' as well as 'sensum' are mental concepts, which does not preclude any physical realisation. It is the concept, not the subject matter or realisation, to which I objected when I said that sensa could not be explanatory items – physical events by themselves have no power to explain anything outside of a conceptual framework.

The following analogy might make clearer what I am claiming, and what not. Consider coins. Suppose that we have both the coins issued by the central bank which are legal tender, and fake coins brought into circulation by counterfeiters. Both the real and the fake coins are (realised by, or identical with) pieces of metal; that much is uncontroversial. But 'piece of metal' is not a *monetary* concept which can *explain* why one would mistake a fake coin for a real one. (It might however be used to explain why something is a fake coin.) Fake coins are fake coins simply because they look like real ones; by referring to something as a fake coin, the need for explaining why someone might be taken in has as it were been pre-empted. The disanalogy with mota lies herein, that 'piece of metal' should be compared to 'bodily movement', not to 'neutral bodily movement'. Both intentional actions and involuntary movements may be realised by bodily movements; but not by neutral bodily movements, because an action cannot be identical with a neutral bodily movement. What the equivalent of mota could be

– a monetary item neutral between real and fake coins, explaining how one could be taken in – is unclear. The lack of such a concept should make us suspicious of mota and sensa, too.

The second possible disanalogy is that mota would not seem to give rise to scepticism like sensa do. The sceptical problem with sensa is that if they are world-independent (intrinsically neutral between genuine perception and delusion) then we end up having no way of telling whether our experiences tell us something about the 'external' world. In fact, there is a precise mirror-image of this scepticism for action: if mota are reason or intention-independent (intrinsically neutral between intentional action and mere bodily movement), then we end up having no way of telling whether a movement is somebody's action. Scepticism about the external world and about other minds (or other agents) are each other's mirror-image, introduced by the same metaphysical arguments.

As the causal theorist employs the term 'neutral bodily movement', they are things that we are supposed to see when we look at people, their actions, and what happens to them. Presently, I will argue that we do not in fact see such things, and that to suppose that we do rests on a dubious epistemological position. For now, I just wanted to show that mota are not any more acceptable than sensa. The discussion of mota could shed light on what, more precisely, is the problem with sensa. Sensa are not objectionable *just* because of being private and inner. It may very well happen that empirically a correlation is established between some brain states of a subject and his visually experiencing something. Such brain states, then, would be present both when the subject is genuinely perceiving and when he is delusorily visually experiencing. But this does not establish the conclusion that the brain states - the existence of which we do not doubt for one moment, and which we know how to recognise with our instruments - must be (token-) identical with things called visual experiences, which have specific explanatory tasks assigned to them within a theory other than neurophysiology.

4.2.5.2 Atomicity and inference

The way in which the causal ancestry claim is set up presupposes a certain questionable epistemological position, which I will call epistemic atomism. According to epistemic atomism, all we ever see are atomistic observation units, which are independent of each other, in the sense that how things are in one unit has no implications of necessity for how they will be in the next. Hume, with his

theory about causation, is a well-known example of a philosopher subscribing to epistemic atomism. He argued that nothing in the events we observe shows that they are related as cause and effect. Our idea of the necessary connection which we call causation arises from the constant conjunction of certain impressions, such as the impressions of fire and heat. From the constant conjunction we infer - mistakenly, according to Hume - that two events are necessarily connected: the necessary connection is not itself something that we could see even if it existed. This is not surprising: since for him atomistic impressions are epistemically basic, a fortiori we cannot observe any relation between the events or objects of our impressions.

While Hume's reasoning about causation is widely accepted, it is worth noting that the epistemic atomism which underlies it is not very plausible, and has some dire consequences. Anscombe points out that by his own strict standards Hume gets himself in trouble:

> "..when we consider what we are allowed to say we do 'find', we have the right to turn the tables on Hume, and say that neither do we perceive bodies, such as billiard balls, approaching one another. When we 'consider the matter with the utmost attention' we find only an impression of travel made by the successive positions of a round white patch in our visual fields...etc. Now a 'Humean' account of causality has to be given in terms of constant conjunction of physical things, events etc., not of experiences of them. If, then, it must be allowed that we 'find' bodies in motion, for example, then what theory of perception can justly disallow the perception of a lot of causality?" (Anscombe 1971, p.68)

If we go along with epistemic atomism, it must be doubted that we can see material objects in the usual sense, viz. objects persisting over time. How can we tell whether the object we see at one instant is identical with that which we see the next instant? Nothing in our momentary impressions tells us that they are impressions of one and the same object. Similarly, we cannot perceive movement. The most we will see are successive 'bodies' at successive positions.

Obviously all this is implausible. We do see persisting material objects, and we do see movement. And we also see the operation of many instances of causal concepts: we see people push shopping trolleys, water wet clothes, matches light cigarettes, the wind move tree branches, balls break windows, and so on.

The causal ancestry theorist about action also subscribes to epistemic atomism. He contends that we do not see people act: we see their bodies move. Then, taking into account the context, we infer either that they acted or that something happened to them. I showed, in section 2.2.2, that this is not how things are: neutral bodily movements, or mota, are not the basic unit of observation. Instead, we see people's actions *as* actions in a context. To say this is just to deny

epistemic atomism: it requires effort to abstract from the context and see only somebody's bodily movements, just as it requires effort to see a player's kick-off and the movement of the football as independent but consecutive events. We have been seduced by the clinical thought-experiments of armchair-philosophy, which point out that it does not make a difference to what we see whether a person raises his arm or his arm rises. When do we ever simply want to raise our arm, blocking from consideration any context in which such an action might take place? Similarly, we normally have no problem in distinguishing the numbers 6 and 9; it is however possible to create a pseudo-problem by writing one of these on the middle of a blank sheet of paper, and then asking whether it is a 6 or a 9 - after all, there is no intrinsic difference!

The causalist about perception also subscribes to a form of epistemic atomism, by assuming that the visual experience is the basic epistemic unit. The visual experience is basic in that we can presumably have the same visual experience whether we genuinely perceive or hallucinate: therefore nothing about the visual experience itself can tell us which of the two is the case. Thus we see that to accept the argument from delusion is to commit oneself to epistemic atomism. Again, it has unwelcome consequences: nothing about our visual experiences will tell us whether they fit the world or not, and thus we can never be sure. Therefore, knowledge about the 'external' world is impossible - even whether it exists must be doubted.

It may be worthwhile to mention the analogous sceptical problem for action here. Just as we can never be sure that our visual experiences fit the world - or even that there exists such a world - because we are as it were unable to reach outside, we can similarly never be sure that people act. According to the causalist we would know for sure that our experiences are of the world if we know them to be caused by it; similarly we know for sure that somebody acts if his bodily movement is caused by his primary reason for that action. Normally we infer from the context whether somebody has acted rather than having something happened to them (i.e. move their bodies), but this does not amount to fail-safe knowledge that they have. In order to have that, we would need to 'look into their minds' to see whether the primary reason caused the bodily movement. But we cannot do that - this is just the problem of knowledge of other minds, the 'inward' equivalent of the 'outward' sceptical problem generated by sense-data. The problem about knowledge of other minds is generated by the fact that if we think that people's mental states are inner causes of their behaviour, then the best we can do to know about their mental states is to compare their behaviour with our own and postulate that, by analogy, the inner causes must be the same as in our

own case. But such induction on the basis of one case - our own - leaves of course plenty of room for the sceptic; and, if we can't know people's mental states, nor can we, on the causalist's construal - ever know whether they act at all.
Of course many philosophers have argued that this anti-Cartesian argument against sense-data need not have grip on visual experiences as required for the argument from delusion. The argument only has grip if sense-data are thought of as epistemic intermediaries, of which we are aware, but to think that visual experiences are like that is to commit the 'sense-datum fallacy'. Visual experiences are simply things we *have*, but we are not aware of them in some perceptual sense. However, I have shown in the previous chapter that this line of reasoning is incompatible with talking of visual experiences as indistinguishable. The 'sense-datum fallacy' is no fallacy at all.
The same reply to scepticism is available here as in the case of action: to think of visual experiences as the basic epistemic unit is a mistake. Our experiences always come within a context, abstractions from which require effort: being well awake and conscious, I see and feel the chair on which I am sitting, rather than inferring that there must be a chair on the basis of my experiences. Consequently, I hallucinate or perceive - and I am usually able to tell which of the two is the case, but not on the basis of a hazardous inference from my having neutral visual experiences. To admit this is to rob the argument from delusion of its bite. The sceptical argument can of course be pushed, but from the argument that we can always be mistaken nothing much follows - at least not that our perception starts with a basic epistemic level which is neutral with respect to how the world is.
If I am right that the causalist about action or perception subscribes to epistemic atomism, then he has to face a further difficulty. On the one hand, because of his epistemological assumptions, he is a natural ally of Hume's regularity theory of causation. However, on the other hand, if all there is to causation is matter-of-fact constant conjunction, it does not seem that we thereby have a link that is robust enough for causal theories of the mind. For example, if somebody consistently hallucinates an apple whenever there is an apple in front of him, how do we distinguish that phenomenon from genuine perception? On the Humean regularity theory, since there is this constant conjunction, de facto the apple causes the visual experience. The causal theorist then has to admit that he cannot make sense of this case as described, although there would seem to be no problem in principle with repeated hallucinations consistent with the facts. If, simply in virtue of their regular consistency with facts, hallucinations would have to be declared to be genuine perceptions, something must be amiss with the causal theory.

4.2.6 Conceptual dependency

In section 2.2.8 it was argued that the notion of action is conceptually prior to causation. The argument was based on the idea that, without the notion of an experiment, we cannot give any more robust sense to causation than constant succession. If this was right, then a similar argument is available for perception. For the notion of an experiment requires not only that we be able to intervene in states of affairs in order to vary experimental conditions; it also requires that we be able to perceive how our actions change those conditions, and of course to perceive the outcome of experiments. Stating the point more strongly, we might say that the notion of intervening itself is already dependent on the notion of perceiving. The point is not the mere epistemological one that without perception there is no way of telling what our actions achieve: but, rather, that achieving any intervention is not possible without perceiving the change constituted by the intervention. Both the concept of perception and the concept of action would then be conceptually prior to the concept of causation.

It is no simple matter to establish which of a 'cluster' of concepts is prior to all the others, if any. The argument hinted at here is in need of more detailed support. However, it does go some way towards showing that claims often made by causalists, namely that the notion of acting requires the causal notion of bringing about changes, or that the notion of perception requires the causal notion of being impacted on by the world, could at least plausibly be inverted. Clarification of the concepts of action and perception by means or in terms of the concept of causation thus becomes a less obvious move.

4.3 Causation

Until now, causal theories of the mind of been discussed without first seeking a clear definition or understanding of the concept of causation. The criticisms made of causal theories do not rely on any specific conception of causation. Given that causation is one of those notions that philosophers keep disagreeing on and being puzzled about, it may seem odd that I have said very little about it until now. Hugh Mellor would object: when he tells us about his reasons for writing a book on causation, he writes

> "One [reason] is that causation is increasingly invoked throughout philosophy: in theories of mind and language, of explanation, of decision making and of action, and of knowledge. Yet it is rarely either obvious or shown that what these theories assume about it is true. All too often crucial but contentious assumptions (e.g. that causation is physical, extensional or law-based, or involves necessity) are backed only by unargued intuitions.

> That is not good enough: the more we need to invoke causation, the more we need a full and fully argued account of it to tell us what it is, what it entails and what it can and cannot do." (Mellor 1995, p.3)

In this section I will, for each of the objections to causal claims about the mental identified above, consider what has to be true of the causal relation in order for them to have real bite. I will then very briefly indicate what the main controversies about causation in the literature are. Possibly, even though there is no consensus over what account should be given of causation, the controversies over causation do not play any role in the objections which I have formulated against causalist claims.

The problem of deviant causal chains, to start with, was that something could cause a state or event intrinsically identical to a mental one in a way other than 'the right one'. Does the judgement about what causes what depend on a theory of causation? That seems unlikely; the causal judgements about (deviant and non-deviant) causal chains are in a sense primitive - they are the kind of judgement which a theory of causation is supposed to accommodate. In other words, the task of a theory of causation would be to give an account unifying all those cases in which according to our intuition there is a causal link. A revisionary theory of causation, i.e. one which happened to exclude the deviant causal chains as not causal after all, is of course not to be excluded. But, firstly, no controversy exists in the debate over deviant causal chains over whether deviant chains are indeed causal. Secondly, few will be prepared to concede their judgements about when two events are causally related as mistaken - unless, perhaps, they were persuaded that the alleged cause and effect were already logically related. But in that case, the 'right' causal chains, not the deviant ones, would be the first candidates for rejection. In other words, if one rejected Davidson's argument that it is not causes and effects, but only their descriptions, which can be logically related, then one thereby makes the causal theory vulnerable to the logical connection argument. The problem of deviant causal chains, I conclude, is not theory-dependent.

The same must be said about the objection based on the absence of any suitable causes for Bs in virtue of which they could become Xs: the problem of arational actions. It has nothing to do with a theory of causation whether a cause of the suitable type is present – where 'suitable' is defined not by any theory of causation but by the causal claim being made about mental explanations (FPCE). One does not, in short, need a philosophical analysis of causation in order to decide whether suitable causes are present.

What about the sceptical problems arising because of the postulation of common-factor items such as sensa and mota? An assumption is made here, but it is the

very same one that get the causal argument off the ground in the first place. There is no sure way of inferring the existence of the external world from one's visual experiences, nor of inferring mindedness from someone's neutral bodily movements, because an event's cause is a factor extrinsic to it, and cause and effect are logically independent. The extrinsicality assumption is needed by causal theories, because their main argument (the argument from illusion) is held to exclude the possibility of a differentiating intrinsic feature.

The argument throwing doubt on whether causation is a concept by means of which we can clarify the concepts of action and perception on the grounds that there may be a conceptual dependency running the other way is itself tentatively putting forward (part of) a theory of causation. Clearly, therefore, it makes an assumption about causation. However, there is no hidden commitment to taking sides on any of the main controversies about causation, more about which later.

The conflict between mental states being causally efficient and at the same time supervening on causally efficient physical states – the problem of mental causation – does harbour an assumption about causation. The assumption is that where there are two causes operative, only one of them can be a complete cause of the ensuing effect. If there are two causes which both have the potential to cause the same effect, and they both occur at the same time, then only one can be 'doing the work'. One cause, in such a case, will pre-empt the other. Three remarks are in order here.

Firstly, 'complete cause' is perhaps not an uncontroversial notion: a short circuit may be called 'the' cause of a fire, even if it could only be efficient in the presence of oxygen and combustible material. In other words, most causes are only partial causes. But suppose that there was both oxygen and combustible material present: it seems that in that case either the short circuit or the lit match was the cause of the fire, but not both. Unless, that is, we can distinguish separate effects (e.g. the fire starting in the metering cupboard and the one starting in the bedroom). The case with the alleged overdetermination by mental and physical causes is not affected by these possible objections: all other present partial causes are held constant, and the effect is one and the same. It should be held in mind that overdetermination is not held to be impossible: the problem of mental causation rests on the assumption that it is unlikely to occur *systematically*. The only assumptions made here about causation, then, is that overdetermination occurs when there are two complete causes, and that such overdetermination does not systematically occur.

Secondly, the problem of mental causation is not really about the efficacy of separate causes but about the efficacy of different properties. Therefore, the

problem does not arise due to any problematic assumption as to how causes should be counted: even if we assume the physical cause and the mental cause to be (token-)identical, there is a question about how both the physical and mental properties can be efficacious.

Thirdly, an additional assumption is made to the effect that there is only one kind of causation. If one were to argue that the mental state and the physical state – or perhaps, the mental and physical properties of the one state – both played causal roles, but of very different kinds, then overdetermination does not arise. In other words, if one admits of a pluralist view of causation, then the problem of mental causation does not arise. But there are two related reasons for a causalist to resist this view, and endorse what I in chapter 1 called the 'no pluralism about causation' (**NPC**) claim. In the first place, some account has to be given about why mental causation still deserves to be called causation. For it would not be entitled to the same robust notion of efficacy as physical causation. And secondly, much of the motivation for causalist claims derives from the idea that the mental must somehow fit in with the natural world order. The claim that a reason causes an action in some different sense from how, e.g., a short-circuit causes a fire would be in tension with that.

The basic question pursued by theories of causation is: for two things to be related as cause and effect, what more needs to be true than that both the cause and the effect were instantiated? Hume, famously, thought that nothing more need be the case than that the cause precedes the effect, and that there is constant conjunction. Various writers have tried to make the concept of causation more robust than that: in particular, many have argued that we need to be realists about causation, in the sense that it is a real relation to be found 'out there' rather than, as Hume seems to suggest, it being a projection of the human mind onto the world. But whatever may come out of that debate, I do not see that my discussion of the problems for causal theories depend on that outcome. Other controversies include the question whether or not causation is a relation; and if so, whether it holds between particulars or universals, between events, facts, or properties; and whether there can be indeterministic causation. An eventually emerging theory of causation may bring trouble for causal theorists of mind, for example by denying that causation is a relation at all, or by denying that it is a natural or mind-independent relation. Mental states and events may turn out not to be things that can cause or be caused. But if this were the kind of insight reached, it would simply be an additional problem for causal claims about the mind.

4.4 Summary and conclusions

In this chapter I have looked at the analogies and disanalogies in the discussions about causal claims about action and perception. Although there are a few disanalogies, the analogies were found to be dominant. Objections to causal claims do not automatically generalise to other mental concepts – sometimes because the causal claims are not the same, and sometimes because an objection is specific to the concept – but there are two reasons why the comparison between mental concepts is useful. The first reason is that it may aid better understanding of what is at issue with certain objections. The second reason is that it generates some kind of checklist of difficulties that are central to various claims about the mental. The risk of oversight of difficulties, or of duplicating discussions over a certain difficulty, is thereby reduced.

5 Residual causalist arguments

We have seen that there are no conclusive arguments in favour of causal theories of action and perception. The argumentation for such claims keep running into trouble. Moreover, in the previous chapter it was shown that the problems for the causal theories of action and perception are of remarkable similarity. It is perhaps not a big surprise that this should be so, given that the structure and motivation for the two theories is largely the same. Reasoning on in the same vein, however, we should come to the conclusion that causal theories of the indicated kinds for other mental faculties, by generalisation, are in an equally difficult situation. All this will be hard to swallow for many philosophers of mind.

It may be protested that I am at this stage a bit quick in drawing such sweeping conclusions. Given that I have not yet thoroughly addressed the two considerations that are perhaps most basic to causalist argumentation, this is quite justified. In this chapter that situation will be remedied. The two residual arguments that the causalist may have up his sleeve are discussed. The first one has to do with counterfactual conditionals that are true in cases of acting and perceiving. Secondly, there is the idea that mental kinds could (be discovered to) be natural kinds.

5.1 Counterfactuals

An important step in causalist arguments – both for **FPCE** and **NCA** claims - is that in the case of intentional action, genuine perception, etc., there are certain counterfactual conditionals which we hold to be true. How does this support the causal claims? And how well does this support withstand scrutiny? These are the questions I will consider below.

5.1.1 Action

An example of a counterfactual related to action is "If I hadn't believed that the roof was leaking, I wouldn't have phoned a roofing company". The truth of this counterfactual is presumably what makes "I phoned the roofing company because I believed that the roof was leaking" a good explanation. This is so because the counterfactual enables us to distinguish between the real reason and a mere rationalisation; for Davidson it is at the same time the argument for thinking of the explanation as causal.

The argument from counterfactuals to an **FPCE**-claim about action (**CACAE**), then, could be formulated as follows:

(A1) In cases of intentional action a counterfactual of the form "Had I not believed B and desired D, I wouldn't have performed action A" is true of conceptual necessity[94]
(A2) The counterfactual in (A1) is made true by the antecedent being causally explanatory of the antecedent
(CAE) It is true of conceptual necessity that the events in the antecedent are causally explanatory of the event in the consequent of the counterfactual.

In the next section we will see that a closely related argument supports the **NCA** claim regarding perception. I will there offer reasons for thinking that the second premise is false. In the remainder of this section I will not challenge the above argument as such, but rather cast doubt on the implicit assumption that the truth of the mentioned counterfactual is all-important to accounting for the strength of reason-explanations of action. I do this by considering the counterfactuals that are actually playing a role in action explanation. My claim will be that action-explanation makes true, besides the type of counterfactual claimed by Davidson, a number of counterfactuals which cannot even be accounted for by a causal explanation and are in fact incompatible with causal explanations.
Here is an example to show which counterfactuals surround the explanation of action. Suppose I dig a hole in the lawn in my garden, because I believe a treasure to be there which I want to retrieve. The counterfactual figuring in **CACAE** would then be, "had I not believed a treasure to be under the lawn, I would not have dug a hole there." It is this one which supposedly supports the idea that the given explanation is a causal explanation. But there are other counterfactuals, importantly made true by the explanation of my action. Here are a few:

> Had my spade broken, I would have borrowed or bought another one.
> Had the soil proved too hard to dig, I would have sharpened the blade of my spade, or rented a mechanical digger.
> Had I not found the treasure where I expected to, I would have dug holes in other places.

And so on: these counterfactuals are about what happens when something goes wrong, or does not yield the expected result. They serve to distinguish the case at hand from ones in which I dig to try out my new spade, or to get some physical exercise, or was just curious about the kind of soil under the lawn.

[94] Complicating the formulation here is that there is no conceptual necessity as to what the contents of B and D are for a given A. A more precise formulation might thus be "In cases of intentional action there is some belief-desire pair B,D such that a counterfactual..(etc.)" For the discussion following here this is of no importance.

Another example is from Arthur Collins[95]. Someone flips the switch because he wants the light on. Not only would he, supposing this to be the right explanation of his action, not have flipped it had he not wanted the light on, but he would probably have done a series of things had the light not come on by flipping the switch: he would have checked the bulb, checked the fuse, checked the wiring, asked the neighbour if his light is working, paid his electricity bill, and so on.

Admittedly there is substantial indeterminacy in which of these counterfactuals are actually implied. Someone might start digging for a treasure in his garden, and give up if at the first stab the spade breaks. Or flip a switch to turn on the light and, finding it does not work, shrug and stay in the dark. Not everybody gives up at the same point or is equally resourceful as anybody else, nor is this required for the reason explanation to be true. But it does seem essential to intentional action and the explanation thereof that at least *some* counterfactuals of this type are true. For my point is that *no* such counterfactuals are made true by causal explanation of an event.

Let me formulate what a counterfactual of this type would look like if it were associated with a genuinely causal explanation. Take, as an example

> The explosion occurred because there was a gas leak

If this explanation were to behave as in the case of action, a range of counterfactuals would have to be true if there were some inhibiting factor for the explosion's occurrence, or in case the consequences were somehow not appropriate to there being a gas leak. What we might mean by the latter is unclear in any case. In the case of reason explanation, a consequence appropriate to (say) somebody's wanting the light on is that the light goes on. But, unlike what is the case with primary reasons, there are no obvious standards of appropriateness here. Were we to assume that there are such standards – for example, it may be an appropriate consequence of a gas leak that a building be destroyed - then counterfactuals like the following would have to be true:

> Had the building not been destroyed by the explosion, then an earthquake would have occurred.

Counterfactuals relating to the former case (overcoming inhibiting factors) would be like this:

> Had the explosion not occurred, then a switch would have produced a spark.

Clearly, such counterfactuals are not made true of events by their figuring in a true causal explanation. Therefore, it must be that causal explanations are unlike action explanations.

---

[95] see Collins 1987.

The **FPCE**-causalist might try the following reply. "Perhaps you have shown that action explanations have characteristics that other causal explanations do not have. However, couldn't it be the case that action explanations are causal *as well as* what we may call 'purposive'?" I have indeed not shown that this cannot be the case. But the causalist's claim was that the fact that action explanations are causal explanations accounts for the force that they have. As we have seen, action explanations have considerably *more* force than accounted for in this way, and that casts serious doubt on the causalist's claim. But have I shown that these additional counterfactuals are *made true by* the action explanation? If they are features of (perhaps a specific type of) action, but are not in any way linked with the action explanation, **FPCE** stands stronger again. However, this is implausible. It is nothing to do with the action of flipping the switch *an sich* which makes that I would check the bulb etc. had the light not come on. I would do all these things because the explanation "I flipped the switch because I wanted the light on" is the right one. Had I flipped the switch because I wanted to exercise my finger, then the first explanation is not true anymore, and a different range of counterfactuals applies. (Had I not managed to flip the switch I would have done other exercises, I would have called the GP, and so on.)

We have thus seen that besides the counterfactual in the argument as I reconstructed it above (**CACAE**) there are a number of other counterfactuals made true, which are important to the explanation of intentional action. The claim that the force of the 'because' is accounted for by the fact that reason explanation is causal explanation acquires a very hollow ring to it. But that is not all: for in the following sections we will see that the causalist's argument from counterfactuals rests on a false premise.

5.1.2 Perception

Turning our attention now to perception, we find a more explicit commitment to counterfactuals. Smith & Jones write

> If things would not look any different to Jack even if the cat were not present, then he can hardly count as seeing the beast in front of him. To see the cat, Jack must be – so to speak – visually locked onto it. Similarly, for it to be the case that Jill hears the bell, then the bell must be causally responsible for her auditory state. If things would not sound any different to Jill even if the bell were left untouched or were completely absent then she cannot count as really hearing it. (Smith & Jones 1986, p.85)

And Grice writes

> ...it is logically conceivable that there should be some method by which an expert could make it look to X as if there were a clock on the shelf on occasions when the shelf was empty...If such treatment were applied to X on an occasion when there actually was a clock on the shelf, and if X's impressions were found to continue unchanged when the clock was removed or its position altered, then I think we should be inclined to say that X did not see the clock which was before his eyes, just because we should regard the clock as playing no part in the origination of his impression. (Grice 1961 p.69)

Note, that in the examples given here we start out in the one case from a point about explanation, in the other from a point about epistemology, but arrive equally at similar counterfactuals. This invites the thought that the basic motivation really lies in the truth of these counterfactuals in the relevant cases.
Let me state explicitly what the causalist argument from counterfactuals in the case of visual perception (**CACP**) says:

(P1) In cases of vision a counterfactual of the form "Had O not been there, S wouldn't have seen it" is true of conceptual necessity
(P2) The counterfactual in (P1) is made true by a causal link between antecedent and consequent.
(CP) It is true of conceptual necessity that there is a causal link between the events described in antecedent and consequent of the counterfactual

From **CP**, **NCA** about perception follows: it is part of the concept of seeing that the subject's state of seeing O be caused by O. It is easy to see that **CACP** is closely related to the earlier argument from counterfactuals in favour of the **FPCE** claim about action.
My strategy in what follows will not be to attack P1, but P2. My attack proceeds in two steps. First, I show that the connection between counterfactuals and causation is not as strong as often assumed or at least suggested. However, this leaves room for the reply that even if I have shown that there are non-causal as well as causal counterfactuals, I have not shown that the relevant *mental* counterfactuals of the sort used by the causalist in his argument are not causal ones. The next step in my objection to the argument from counterfactuals will address that point.

5.1.3 Counterfactuals and causality

Counterfactual conditionals are traditionally associated, in one way or another, with causal links. Lewis[96], for example, gives an account of what it is for there to be a causal link between event A and event B with the help of a counterfactual conditional. Others, on the other hand, make use of causal relations in trying to formulate truth-conditions for counterfactual conditionals.

Is the step from counterfactual to causal link a valid one? Let us take a closer look at the following biconditional:

A caused B <=> If A hadn't happened, B would not have happened.

The left side of this biconditional I will call the causal statement (CS); the right side I will call the counterfactual conditional (CC). If CS => CC is true, then a causal link is sufficient for making the associated counterfactual conditional true; and if CC => CS is true, then a causal link is necessary for the truth of the counterfactual conditional. Of course, it is the necessity claim which is of primary interest here; its truth would validate the causalist's argumentative step. But the sufficiency claim should not be lost from sight; should the necessity claim prove false but the sufficiency claim true, the question would be invited "but what else than a causal link can make the counterfactual conditional true?" In other words, the causalist's case would be supported by the sufficiency claim, albeit in a less than conclusive manner.

5.1.3.1 Does a causal relation imply a counterfactual?

At first sight the sufficiency claim, CS => CC, seems very unproblematic. Because intuitively our first reaction to the question "what does it mean when we say that A caused B?" is "it means that if A hadn't happened, B wouldn't have happened." It is, however, not without problems. Take the case of causal overdetermination. C_1 causes E, C_2 causes E, and C_1 and C_2 are both present. In this case the counterfactual conditional " If C_1 hadn't happened, E wouldn't have happened" seems to be false; for if C_1 hadn't happened, C_2 instead would have caused E - so E would have happened after all. So we have a situation in which CS is true but CC isn't, and therefore the sufficiency claim fails to be true. Overdetermination is not the only problem: there is also the kind of case in which one cause (C_1) prevents the effectiveness of another cause (C_2), but by itself brings about the same effect; or in which there is some process C bringing about either C_1 or C_2, these in turn causing E. I regard these cases as variations on the same theme.

---

[96] Lewis 1973

Cross[97] tries to rehabilitate the connection between counterfactuals and (event) causation. In short, his point is this: in order to restore the link we should restrict ourselves to a set of test worlds for the counterfactual, namely those in which it is true that if the original cause had not occurred, no other event instead would have caused the effect. But this proposal will not do. What Cross suggests here is an ad hoc way of making any counterfactual true: find out by which possible worlds your counterfactual is made untrue, then restrict yourself so as not to test the counterfactual in exactly those possible worlds. The counterfactual "If I hadn't kicked the cat this morning, I wouldn't have had this accident" is of course true when restricted to possible worlds in which the counterfactual "If I hadn't kicked the cat this morning, no other event would have caused my accident." is true. In other words, in this way the counterfactual is easily made true even if there is no causal connection at all. With this construction causal statements, even in cases of overdetermination, are made to imply counterfactuals again; but this is accomplished at the price of 'true' counterfactuals being implied by untrue causal statements as well. This does not improve the health of the sufficiency claim.
Our problem, then, is not solved. But apart from saying that the sufficiency claim is not unproblematic I don't want to dwell on this point, and move on to a consideration of the necessity claim, arising from reading the biconditional in the other direction: CC => CS.

5.1.3.2 Does a counterfactual imply a causal relation?

Does the truth of a counterfactual warrant the inference that the events thus related also figure in a true causal statement? After some reflection we must conclude that this can't be a valid inference, since there is a number of counterexamples - a counterfactual can be made true by lots of other facts besides a causal relation. This is pointed out by Kim[98]. For example : "If I hadn't parked my car on the double lines, I would not have broken the law." We would in suitable situations (i.e., those in which parking on double yellow lines constitutes a parking offence) agree with the truth of this conditional; but I doubt that without some dubious artificial stretching of the meaning of the word 'cause' we would agree that "My parking the car on the double yellow lines caused me to break the law." This is because my parking etc. would seem to be the same event as my breaking the law, picked out under different descriptions: my breaking the law consists in parking such and so. We might think we can save the situation by

[97] Cross 1992
[98] Kim 1973

giving some restrictive account of event individuation, but that won't get us anywhere. Consider the explanation: "I became an uncle because my sister gave birth", and the associated counterfactual "If my sister had not given birth, I would not have become an uncle." If one wants to argue that we simply have here two different descriptions of the same event, it does not help: no event can cause itself, and therefore the 'because' cannot be a causal one. But if, on the other hand, one should argue that the explanation does mention two separate events, there is the problem that insistence on a causal interpretation of the 'because' commits one to the possibility of instantaneous causation at a distance: I become an uncle the very moment my sister gives birth, even if she is at the other side of the world.

My sister's giving birth is not a cause of my becoming an uncle; I *thereby* become an uncle. In the same vein, my parking on the double yellow lines *is* breaking the law. These counterfactuals are true by convention: the definition of what is needed in order to be(come) an uncle, or what counts as breaking the law. Obviously, there exist non-causal counterfactuals. But of what type are the counterfactuals figuring in the causalist's argument?

5.1.4 Mental counterfactuals

In the previous section I have shown that there are causal and non-causal counterfactuals. In order to continue my objection against the argument from counterfactuals, I will have to say a little bit more about the distinction between causal and what I will call conceptual counterfactuals. How do we settle the question of what kind a particular counterfactual is?

Consider, as an example, the following two counterfactuals:

(1) Had the road not been slippery, then the car would not have skidded.

(2) Had I not parked on the double yellow lines, then I would not have broken the law.

I think that the best way of bringing out the difference between the two is David Owens' notion of empirical content[99]. He introduces this notion in order to be able to distinguish between causal and non-causal explanations. The idea is that causal explanations are those explanations which are supported by generalisations with empirical content. A generalisation with empirical content is one which is subject to tests independent of the specific case referred to in the explanation[100]. I will, in what follows, assume that we can substitute

[99] see Owens 1992, ch.4

[100] ibid. p.72

'explanation' with 'counterfactual'. Not every counterfactual has an associated explanation, but on the other hand every explanation does have an associated counterfactual. I will elaborate later on whether making use of the notion of empirical content yields the concept of 'causal counterfactual' which we need in the current context.

To establish the truth of counterfactual (1) we would normally undertake certain empirical investigations. We might look at the correlation between roads being slippery and cars skidding off them, or investigate the physics of skidding, or verify that the car's steering installation was functioning as it should. Such investigations all require that we investigate one or more individual cases, and from there make an 'inductive leap' towards a generalisation that would support the counterfactual. These are generalisations such as the general "cars tend to skid on slippery roads", and more specific ones such as "cars with velocity v and friction f between road surface and tyres skid". However, the truth of counterfactual (2) is established in a different way: from the comfort of our armchair we verify the Highway Code for what it says about parking on double lines. No correlations, or investigating the physics of law-breaking or double lines, are relevant here. It makes no sense to park on double lines at different locations, or to park at the same location with another car in order to check whether in that case I am breaking the law too. This conditional is, as Owens puts it, "supported by a law of the state rather than a law of nature. A law of state is not an empirical generalisation, it is not refuted by crime – hence the need for a police force."[101]

The distinction between empirical and non-empirical content can be a rather subtle one. For instance, "had I not parked on these double yellow lines, then I wouldn't have got a parking ticket", despite its similarity to (2), *is* a causal counterfactual: the extent to which there turns out to be a correlation between people parking in certain places and their receiving parking tickets does tell us something about the truth of this counterfactual. The relevant generalisation is open to refutation if, for example, traffic wardens pass here only on Tuesdays, or happen to have decided only to give tickets to Fiats.

A possible difficulty for Owens' account are counterfactuals like this: "If the mean molecular kinetic energy in my porridge hadn't been so high, it wouldn't have been so hot". If we follow Kripke, then this is a necessary a posteriori truth. So it would seem that there is empirical content here, but we should probably want to deny that the mean molecular kinetic energy causes the porridge to be hot, unless we interpret the latter as the porridge *feeling* hot. Owens does defend

[101] ibid. p.72

his account against this objection; however, for current purposes, what we need is a *necessary* condition for a counterfactual's being causal, not a *sufficient* condition. Therefore this type of objection is irrelevant to us here.

Now here is my objection to the causalist, illustrated for the case of the causal theory of perception (**NCA**) and argument **CACP**. If we compare the mental counterfactuals with causal counterfactuals, then it becomes clear that the former do not possess empirical content. Take "Had the platypus not been there, then Pia would not have seen it." There is a difference in the empirical verification that we can do between this counterfactual and the obviously causal "had the road not been slippery, then the car would not have skidded." Let us look at the relevant generalisations in the case of the skidding car:

a) Cars tend to skid on slippery roads
b) Cars may skid due to high speeds (or bald tyres; brake failures; etc.)
c) Normally cars don't skid.

The empirical content of the counterfactual comes from the fact that all these three are inductive generalisations, which can be verified in the usual empirical manner. The counterfactual says that in *this* case the skidding was due to slipperiness, not to high speeds or bald tyres. (There is normativity involved here; it is plausible that a car driven at very low speeds with special tyres and an ABS braking system will not skid, no matter how slippery the road. The term 'normally', in generalisation c. gives this away. However, whereas the norms we use in the end determine our judgement on the truth of the counterfactual, this does not mean that empirical investigation becomes irrelevant.) The empirical content of the counterfactual has to do both with the inductive generalisations, and the question under which of the generalisations the case under investigation falls (how fast was it actually going, what was the mechanical condition of the car, and what was the road condition?).

Now let us look at the perceptual counterfactual. There are the following relevant generalisations, analogous to the case of the skidding car:

d) people tend to see things in their direct field of vision
e) people may see things reflected in mirrors, etc.
f) people don't see things not in their direct field of vision nor reflected in mirrors etc.

Analogous to the other case, the counterfactual says that Pia's seeing the platypus is a case of its being in her field of vision, and not because of a mirror, periscope or whatever. But now notice the differences. Firstly, f. clearly is no inductive generalisation (nor are d. and e., but less obviously so). What would disconfirm it

is a case of somebody seeing something not in their direct field of vision (nor etc.). But if such a case were to present itself, would we not simply deny, either that they saw it, or that it was outside their field of vision (nor reflected etc.)? In other words, generalisation f. is true in virtue of the concepts used.

Secondly, and relatedly, whether Pia's case is one in which the platypus is in her field of vision – that is, the question whether the generalisations apply – is something which we normally establish by asking her whether she sees it. Comparing it with the skidding car case, this is like establishing whether the road is slippery by verifying in some indirect manner (we cannot verify *directly* what Pia sees, we can only *ask* her) whether the car skids. So whether the platypus is in her field of vision (the most likely meaning of "it's being there") is something we decide on the basis of whether she sees it, and that we have only indirect access to. This point about indirect access is far from trivial, since as we have seen much of the discussion about perception revolves around the distinction between seeming to see and genuinely seeing. **CACP** is meant to deliver a conclusion about the concept of seeing, not of seeming or reporting to see. These disanalogies show that the perceptual counterfactual has no empirical content, and thus that it is not causal.

I said something about how we 'normally establish' whether something is in someone's field of vision. But is Pia's seeing the platypus more than prima facie evidence that the beast is in her field of vision, which could be refuted by showing that there is no causal contact between Pia and the platypus? Whether it is anything more than prima facie evidence is the very question that is at issue; if there were an independent way of testing whether it is in her field of vision, the counterfactual would have empirical content. But notice that if we are to assume that the absence of causal contact precludes something from being within Pia's field of vision then we are begging the question in favour of a causal ancestry account of perception. We can, of course, define the notion 'field of vision' in a causal way, but then the empirical content of the counterfactual and hence the causal ancestry condition on 'perception' get introduced *by definition*, which is something else than the causalist claims to be doing. Our ordinary ways of talking about perception can, and do, go without such a definition.

It might be thought that I introduce the dependency of verification of antecedent and consequent – for it is that which creates the disanalogy, and the lack of empirical content - by a particular translation of the antecedent. Why should we not take the vague expression 'a platypus is there' to mean 'a platypus exists and is located at coordinates (x,y,z)?' But this will not work. The perceptual counterfactual is not to be understood as saying that, had the platypus moved 1

inch to the left, Pia would not have seen it. Specifying an area without reference to the concept of field of vision is not possible either: for where, then, are the boundaries to lie? However, could there not be normativity – or perhaps we should say, vagueness - inherent in 'there', just like there was in 'slippery'? Nonetheless, there will still be clear cases in which no platypus is "there", just as there will be clear cases in which the road is not slippery. Therefore the awkward cases, in which somebody reports seeing a platypus although none is there, will not go away.

To reduce the disanalogy between the counterfactuals to its bare bones: what has been shown is simply that if somebody reports seeing a platypus when none is there, the truth of the generalisation supporting the perceptual counterfactual will prevail over the assertion that the person saw the platypus – in other words, we conclude that after all he didn't see such a beast. In the case of the skidding car, by contrast, there is nothing strange about a car skidding even though the road is not slippery: the fact that the road is not slippery is no reason whatsoever to doubt that the car skidded. (But again, by contrast, not having parked on double yellow lines is a reason for assuming you haven't broken the law – unless there is some other traffic sign, which is analogous to the case in which Pia sees the platypus in a mirror.) This disanalogy between the two counterfactuals should be enough to call into question whether the perceptual counterfactual is causal, even before the difference is analysed in terms of empirical content, as I have done.

The difference between the perceptual counterfactual and the causal one should, on reflection, not surprise us. It is the consequence of the fact that the perceptual counterfactual is a necessary truth, a truth, that is, in virtue of how we use the concept of seeing. Or in yet different words: the perceptual counterfactual is grounded in the concepts used, rather than being grounded in a causal relation, which grounding in turn, according to the causalist, would tell us something about the concepts. This is not the case with the skidding car-counterfactual: we are perfectly willing to consider cases in which the car skids in circumstances where the road is not slippery, or doesn't where it is. 'Skidding' is a concept we use in such a way that instances of it may be due to any of a number of factors.

Putting the point in this way invites the following reply from the causalist: "You have just given a version of the old logical connection argument! But whereas you may have shown that a normal causal counterfactual is different from a conceptual counterfactual, and that there are reasons to believe that the perceptual counterfactual in significant aspects resembles the latter, you haven't

shown that a counterfactual could not, in your sense, be both conceptual *and* causal."

This reply to my objection, though plausible at first blush, will not do. The point emerging from the discussion of the logical connection argument in section 2.2.3 was that Davidson has shown us that there is no contradiction in two events being causally related and their descriptions at the same time logically related. However, we are dealing with something quite different here. The argument is not over whether the events in antecedent and consequent *can* be causally related, or whether they, as a matter of fact, *are* so related. I have not denied the claim that causal processes are involved when somebody perceives something (**CPI**). I did introduce non-causal counterfactuals by giving examples in which events in antecedent and consequent cannot be causally related. However, this should not be understood as saying that all, and only, those counterfactuals in which there cannot be such a causal link, are non-causal counterfactuals. My argument with the causalist is about the *type* of counterfactual conditional figuring in **CACP**. For **CACP** to succeed, it must be the case that the counterfactual is *made true by* the causal link between the two events. That is a matter dependent on the descriptions used for the events; I have tried to capture it with the notion of empirical content. Not only is it *not settled* by whether antecedential and consequential event are causally related, but that is *irrelevant* to the question. A counterfactual can be devoid of empirical content even if antecedential and consequential event are causally linked[102]. In a variation on Davidson's well-known example[103], one may think of the counterfactual "if the cause of B hadn't happened, B wouldn't have happened".

To recapitulate my objection to the causalist argument from counterfactuals: for the argument to succeed, the counterfactual must be causal, i.e., have empirical content. This is something else than that the events in antecedent and consequent are causally related: the counterfactual must be made true by the obtaining of the causal relation. Comparing the kind of empirical investigations we might undertake to support our counterfactual in an obviously causal case and in **CACP**, we have to conclude that the latter is not causal. The causalist's argument therefore does not succeed. My objection does not fall prey to Davidson's answer to the logical connection argument. I will not go through the moves of formulating this same objection for **CACAE**: it proceeds along similar lines[104]. It was, anyway, shown that the **FPCE** claim about action is not so powerful as it

[102] The converse is also possible. That is needed for the **CACAE** argument (counterf. argument for causal action explanation), if one wants to make an **FPCE** claim about action while denying the **NCA** claim, as the practical realist does. (see chapter 2).

[103] Davidson 1980 p.14

[104] although there is a complicating factor; see footnote 94

may seem: there are many true, and importantly relevant, counterfactuals in any given case of intentional action the force of which is not accounted for by conceiving of action explanation as causal explanation. Short of being able to account for such counterfactuals, the causal explanatory claim is considerably weakened.

5.2 Empirical evidence, natural kinds and essentialism

I have earlier on argued that it is not of any use to a causalist to appeal to findings of empirical science. Science can only reveal *contingent* truths, and therefore doesn't bring necessary *conceptual* truths within reach. However, the causalist might make an appeal to the notion of a natural kind. The idea would be that mental kinds are, or are discovered to be, natural kinds. But we have to be careful as to which claim may be available on this basis. In the following, I will first explain the notion of a natural kind, and consider whether mental kinds could be natural kinds. Then, I will make explicit a causalist argument about mental concepts which is based on natural kinds. As we shall see, a causalist claim about mental concepts of type **N(S)CA** will not gain any new support. However, there is room for a weaker causalist claim of type **CE**, which says that even though this is not reflected in the concepts themselves, the subject matter that concepts of mental states refer to is such that it belongs to their essence that mental states are causally active and acted upon. This essentialist claim would be implied by the stronger conceptual claim which I have been arguing against; I will consider whether it can stand on its own. This assumes a distinction between natural kinds and natural kind concepts. I will argue that if the essentialist claim can do without the conceptual, I have no quibble with it. My main target is the conceptual claim, the arguments against which remain unaffected by the appeal to natural kinds.

5.2.1 Natural kinds

What is a natural kind? Kripke's view of natural kinds[105] is as follows. Take water as an example. Water is a substance with all kinds of indicator properties – liquidity, transparency, tastelessness, thirst-quenchingness etc. It has a hidden structure that is essential to it: namely, the molecular structure represented by the chemical formula H_2O. This hidden molecular structure explains all of the indicator properties: water's being H_2O explains why it is liquid, transparent, and

[105] Kripke 1980

so on. At some point in time it was discovered that the stuff we called water all along is in fact H_2O. By saying that water is a natural kind we mean that water is a substance such that if any, or even all, of the indicator properties were absent, it would still be water as long as it has the right chemical formula. The converse is also true: all the indicator properties could be present, and yet the stuff at hand *not* be water – namely, if it does not have the required molecular formula. Now, 'water' and 'H_2O' are both rigid designators. Rigid designators are names which in any possible world have the same referent. ('Descartes', for example, is a rigid designator, whereas 'the inventor of the paperclip' is not. The first one refers across different possible worlds, i.e. necessarily, to the same individual, but not the latter. Somebody else might have invented the paperclip, but Descartes might not have been somebody else, since by that token he would cease being Descartes.) The identity water = H_2O holds therefore of necessity, and because we have discovered this identity, it is an a posteriori necessity.

So, on Kripke's conception, water's being a natural kind is a matter of the obtaining of the a posteriori necessary truth of the type 'water = H_2O'. What does it mean to say that this is a necessary truth? It means that if we were to conceive of a tasteless, transparent (etc) liquid with a molecular formula other than H_2O, we would thereby not have conceived of water. Strictly speaking, then, the assertion "It might have turned out (i.e. in another possible world we might have made the discovery) that water was not H_2O" is false. This seems counterintuitive; but this possible-world counterpart of water would not have been water, by virtue of being something other than H_2O. Kripke does however, in order to save intuitions regarding discoveries that might have been made, allow that something pretty close to this might be true: namely, "a liquid with all the same phenomenal qualities that water has might turn out to be not H_2O"[106]. According to Kripke, the truth of the latter assertion accounts for the strong intuition behind the former assertion.

On this story of natural kinds, it is important that something is a natural kind whether or not we have discovered this to be so. We could pick out water by means of the indicator properties before we knew of its hidden structure. However, the discovery of the molecular structure enables us to settle disputes over 'problematic' instances of water: by looking at the molecular structure we can settle the question whether ice, vapour, and perhaps certain transparent tasteless liquids, are instances of water or not.

In explaining what a natural kind is I have not yet said anything about which kinds of thing are natural kinds. If we are to have any causalist argument,

[106] See Kripke 1980 p. 142, p.151

presumably mental events or states need to be natural kinds. Are mental kinds natural kinds? Let us start by considering an example discussed by Kripke[107]: suppose that pain is an excitation of C-fibres. He uses the example to set up an argument against mind-brain identity, in the following way. If the identity of pain with C-fibre excitation is true, it must be a necessary identity, since both 'pain' and 'C-fibre excitation' are rigid designators. But according to the mind-brain identity theory it is a contingent identity. So mind-brain-identity is not true, since it requires that pain might have turned out to be something else than C-fibre excitation. Nor can that statement be true in the sense that something with all the indicator properties of pain (but not the thing we call 'pain') might have turned out to be something else than C-fibre excitation. Anything to which we stand in the same epistemic relation as to pain simply is pain: whatever hurts, is pain.

Kripke's rejection of type-identity of mind and brain thus proceeds on the basis that pain is not a natural kind. Another familiar rejection of the required type-identity between mental and physical states can be found in the argument from multiple realisation. The guiding idea here is that a given type of mental state, such as the belief that this is the last year of the millennium, may be realised differently in the brains of two different individuals, or even in the same individual at different times. This thesis of multiple realisation supposedly derives its plausibility from empirical considerations. But this is not so clear. Our inability to establish type-identities says something about our readiness, in the face of empirical evidence, to abandon our conviction that something is a mental state of a specified type. The argument from multiple realisation therefore turns on the same intuition as Kripke's argument. According to both lines of thought, we are not prepared to allow our conviction that we have an instance of the required type of mental state to be shaken by the discovery that the right type of physical state is not instantiated.

It would be too quick to conclude from this discussion that a rejection of type-identity – currently the dominant position in philosophy of mind - would commit one to hold that mental kinds are not natural kinds. Firstly, even if we suppose the argument, which on the basis that mental kinds are not natural kinds concludes that type-identity of mind and brain is false, to be valid, the converse inference does not follow. However, even if the rejection of type-identity does not require that one accepts that mental kinds not be natural kinds, the fact remains that most philosophers reject type-identity because they believe mental states to be multiply realisable. If I am right, that line of reasoning is based on the idea that identifying something as a mental state on grounds other than hidden

[107] Kripke 1980

physical structure always takes precedence. Therefore, the argument generally used for arguing against type-identity seems to conflict with the idea that mental kinds are natural kinds, even if the position argued for may be compatible with it. Secondly, however, the multiple realisation thesis only excludes that mental kinds be natural kinds *as defined by a specific type of hidden structure*. There is perhaps no necessary a posteriori identity of pains with C-fibre excitation. But we cannot thereby exclude that pains may share some *other* kind of hidden structure, at another level of description. Talk of natural kinds is not so much concerned with reducing phenomena to a basic level of description, as with discovering and stating useful laws and generalisations. To see this, consider again the case of water. Received wisdom has it that water is a natural kind, the molecular structure of which is H_2O. There are, however, two *kinds* of water. The atoms composing the molecules have turned out to have an internal structure which can vary. Most hydrogen (H) has 1 neutron and 1 proton, but some has 2 neutrons and 1 proton. This isotope (it occupies the same place in the periodical table of elements) is also called deuterium, and the water composed of it 'heavy water'. For most purposes it has the same properties as normal water. However, in nature it is very rare, it boils at a slightly different temperature, and behaves very differently in nuclear reactions. Has the discovery that there is heavy as well as normal water shown that water is not a natural kind? No: all water is (necessarily) H_2O. The example does show, however, that between instances of a natural kind there can be differences in hidden structure at a deeper level. This should, on reflection, not be a surprise at all. Take tigers. They are a natural kind. But there are different kinds of tigers: the Bengali tiger, the Sumatran tiger and the Siberian tiger have different genetic make-ups. That we treat them as a natural kind just means that we find that there are enough useful law(like generalisation)s in which all tigers figure. Or all instances of water. Natural kinds and laws go hand-in-hand. We can have laws at different levels – about water, and about hydrogen nuclei -; therefore we can have natural kinds at different levels, too.

Even if laws in the physical domain are unavailable to shape mental kinds into natural kinds, there might nevertheless be mental laws with corresponding natural kinds. But are there any psychological laws? Donald Davidson argues[108] that there are not, and many philosophers agree with him. The best we can do is to formulate all kinds of generalisations, but these can never be exceptionless. This is thought to be due to the normative character of the mental. Be this as it may, the idea that this sharply distinguishes mental laws or generalisations from

[108] Davidson 1980 pp.207-227; see also Heil & Mele 1993

physical laws is based on the mistaken assumption that in physics, exceptionless laws exist. For example, water does not always boil at 100°C: this depends on the atmospheric pressure. Light does not always travel along straight lines: a large mass may deflect it. And so on. All physical laws are *ceteris paribus* laws: there is always an implicit assumption that other factors are held constant, and no matter how many of these factors are made explicit, other ones can always be found that influence the phenomena, and thus 'violate the law'. Therefore it seems reasonable to say that if we allow ourselves natural kinds in physics, we may also do in psychology: requiring exceptionless laws is simply a too stringent demand.

The upshot of the discussion so far is that mental kinds may be natural kinds, although some care needs to be exercised if such a claim is to be held together with an endorsement of the thesis of multiple realisation of the mental. We now need to look at what the causalist's argument is.

5.2.2 The conceptual claim

One claim that may be argued for using the idea of natural kinds is a conceptual claim of type **N(S)CA**. A suggestion to this effect can be found in Grice's paper on perception, in a passage that I mentioned before in the chapter on perception (p.95). He writes

> "..for an object to be perceived by X, it is sufficient that it should be causally involved in the generation of some sense-impression by X in the kind of way in which, for example, when I look at my hand in a good light, my hand is causally responsible for its looking to me as if there were a hand before me, or in which ...(and so on), whatever that kind of way may be." (Grice 1961)

The argument in this passage, designed to deal with the problem of deviant causal chains, seems to be that the causal chain whereby a perceptual object causes a perceptual experience is a natural kind. This would justify confidence that non-deviant causal chains can be distinguished from deviant ones, even if the common hidden structure of such chains have not yet been revealed by scientific discovery. Extending this idea slightly and generalising it, the causalist claim might be that mental states or events of kind X (action, perception, etc.) form a natural kind; and that the hidden common structure consists of, or includes, the state or event having a specified causal ancestry. What this ancestry is exactly would be something for neurophysiology to discover.

We may consider this argument to be a sophistication of the argument from science. According to that argument, we would be ignoring empirical evidence

about how experiences and bodily movements are caused if we held that it is not essential to mental concepts like action and perception that their instances have a specific causal ancestry. My objection was that such empirical evidence can only be evidence for contingent truths, whereas **N(S)CA**, the claim that I am interested in, is a conceptual claim which, if true, is necessarily true. By drawing on the idea of a natural kind, the current argument goes a step further. Empirical discoveries about mental states and events, insofar as they are discoveries about what the common hidden structure is[109], can lead us to necessary truths about mental states and events if we suppose them to form natural kinds. Obviously the assumption that mental kinds are natural kinds plays a crucial role in answering the objection to this argument's predecessor, the argument from science. It does not suffice for the causalist to show that mental kinds could be natural kinds (the question discussed in the previous section): it has to be shown that they are natural kinds, and moreover natural kinds at the level of description which the causalist needs for his argument towards claim **N(S)CA**.

What I have given is not more than the outline of an argument for the causalist, with an indication of which questions need to be answered in any case if it is to work. Below I will consider some of these questions, which have to do with the discovery of natural kinds, and the distinction between natural kinds and natural kind substances. To show that there might be a problem here for the causalist, I offer the following worry. The causalist claim is that a certain causal ancestry is part of the everyday concept of action or perception. However, the argument from natural kinds described above has it that the causal ancestry is part of, or constitutes, the common *hidden structure* of instances of the concept. If that is so, then we can apparently use the concept blissfully unaware of the condition on causal ancestry, just as we can use the concept of water without knowing that water is (necessarily) H_2O. There is a clear tension between this consequence of the natural kind argument and the claim that the causal ancestry condition is part of the everyday concept. It will not do to retreat to the Gricean claim that all the natural kinds argument does is to offer a way of filling in what the causal chain is exactly; for then we are left wanting an argument to establish that the causal chain was essentially part of the concept in the first place.

[109] the proviso is needed: one may make the empirical discovery that some instances of water are transparent, but that is not a necessary truth – think of vapour.

5.2.3 Natural kinds and natural kind concepts

The worry that I outlined could also be reformulated in the following way. "Perhaps perception is a natural kind. But that is a claim about what the subject-matter of the concept 'perception' is like, not a claim about the concept as we use it to refer to that subject-matter. Nothing about the latter follows from the argument from natural kinds." This thought relies on the possibility of using natural kind concepts to refer to non-natural kind subject matter and vice versa. It can be made plausible in the following way.

Take, again, the example of water. As it is in the actual world, this is a natural kind substance to which we refer by means of a natural kind concept. But we might envisage the following two alternative situations, or possible worlds if you like. In the first, water is a natural kind substance, but we refer to it by means of a criterial concept. This means that the concept 'water' refers to whatever has (enough of) the superficial indicator properties (tastelessness etc.). How do we establish that this is the kind of concept we use? Well, suppose we had a sample of something we called water, which upon investigation turned out to have the molecular formula XYZ instead of H_2O. If we react to this discovery by reaffirming that this sample is indeed water, then our concept of water is criterial. Note, that we should not describe that situation as one in which we engaged in some mistaken conceptual practice. It is just that in that possible world the concept of 'water' is a different concept than the one we actually use; let us call it 'water*'. The difference between these two concepts, if they were used alongside each other, would only manifest itself in case a liquid with enough of the indicator properties and a formula differing from H_2O existed, or in case a sample of H_2O for some reason lacked so many indicator properties that it would not be an instance of the concept of water* whereas evidently it would be an instance of the concept of water. Such an occasion need not arise.

The second situation is one in which our concept of water is a natural kind concept – i.e., it is 'water' not 'water*', as explained above – but the substance to which we mean to refer by this concept does not have a single hidden structure. Unlike what is the case in the actual world, some samples of water are H_2O, some are XYZ, others have perhaps yet other molecular formulae. But like in the actual world, we accept that a specialist who can determine what the common hidden structure is and whether a given sample has that structure has the last word in settling the question whether a given sample is an instance of the concept.

Suppose that such a specialist discovers that some of the stuff we have been referring to as water is XYZ instead of H_2O. There are three possible ways of

dealing with this discovery. The first is that we reaffirm water to be a natural kind, comprising both all the samples of both XYZ and H_2O. We would have to acknowledge that the hidden structure we had ascribed to our natural kind was not at the right level; in other words, that we were wrong about the type of natural kind. (Both tigers and lions form natural kinds, but not the same one; they both belong to the natural kind of mammals, though.) This way of dealing with the situation is most suitable if most of the generalisations which were true of H_2O are also true of XYZ.

The second way of dealing with the discovery is to keep the reference constant, and change the concept by which we refer to it: to start using 'water*' instead of 'water'.

The third way is to keep using the same concept, and change the set of referents: we'd have to acknowledge of certain samples that we used to think that they were water, but that we now know better – let's say we call it twater to distinguish it from water.

There is no determinate answer as to how to choose between these three possibilities when confronted with the discovery. Considerations which would be taken into account in making this choice would have to do with the various uses that we have to date been making of our concepts, and the generalisations and laws in which they figured. What is the smallest change, what the most useful choice?

Given the above descriptions of the two situations there seems to be no problem in natural kind substances and natural kind concepts 'coming apart'. This is not altogether true, because the second described situation is not stable. If we were to discover that what we mean to refer to by a natural kind concept does not have the common hidden structure that we thought it had, then we would have to revise our conceptual practices. We cannot consistently use a natural kind concept to refer to a mixed bag of stuffs lacking a common hidden structure. In other words, the only way in which we can use a natural kind concept to refer to a non-natural kind stuff is by persisting with erroneous practices.

In the first situation, the referents of 'water*' and 'the liquid with molecular formula H_2O" are not generally the same. That situation is then not accurately described as one in which we use a criterial concept to refer to a natural kind substance. Some samples of H_2O may not be water*, and some water* may not be H_2O. The overlap between the referents does not justify describing this situation in the way explained. To be accurate, we should say: "Many (or most) of the samples liquid with the hidden structure H_2O can be referred to as 'water*', however, something's being referred to as water* is neither necessary

nor sufficient for its being H_2O." What of the possibility, however, that all existing instances which may be correctly picked out by 'water*' happen to be water, i.e. H_2O? Isn't it possible that contingently, as a cosmic accident, there are no stuffs other than H_2O, correctly picked out by 'water*'? It would be a situation in which any stuff which had enough of the superficial indicator properties happened to be H_2O; in other words, there would not be any close look-a-likes. Though unlikely, it would seem to be possible.

If the above story is right, then the argument from natural kinds is indeed vulnerable to the objection that even if mental kinds were natural kinds nothing would follow from that for mental concepts. If the subject-matter of a mental concept were discovered to form a natural kind this would be some kind of cosmic accident, but it would not require us to start using a natural kind concept. Therefore it is not enough that mental kinds are, or are discovered to be, natural kinds: an argument is needed to establish that mental concepts are (discovered to be) natural kind concepts. For a claim of type **N(S)CA** the causalist would need to establish something like this: "the concept 'perception' is the concept by which we refer to mental states (visual experiences) which have in common that they are caused by the objects that they are experiences of." Or in other words: "the concept 'perception' is that of a mental state / visual experience which has to be caused by the object that it is of." We are now straight back into the arena in which the discussion of previous chapters took place, and in which we have found no such argument to withstand scrutiny. The new thought brought in by the natural-kind argument for causalism, namely that the causalist would be able to get support from empirical discoveries in the natural sciences, turns out to have been blocked by the possibility of natural kinds being referred to by non-natural kind concepts.

5.2.4 Empirical discovery and revision of concepts

The story does not quite end here for the conceptual claim on the basis of the argument from natural kinds. I insisted that the discovery of a common hidden structure does not *require* us to discard a criterial concept start using a natural kind concept. It may however be *useful* to do so: for it would enable us to capture and take advantage of the regularities that nature presents to us on a silver platter. In other words, it might be argued that such a discovery might be a good reason to start using a natural kind concept, and then the causalist could be allowed his conceptual claim after all.

A simple objection to this line of thought would be that the discussion over the conceptual claim **N(S)CA** was supposed to be about the mental concepts we use *now*, and therefore that speculation about how we might want to change them in the light of future discoveries is irrelevant. One might reply that 'water' already was a natural kind concept before the discovery that water has the molecular structure H_2O, but that does not engage with the point. 'Water' already was a natural kind concept because it figured in law-like generalisations, on the assumption that there was such a common structure. In other words, it was not discovered *that* there was a common structure, but the precise nature of it was discovered. The mental concepts that we currently use, by contrast, are not natural kind concepts, at least not of the type which the causalist needs. That, again, is the conceptual point which I hope to have made plausible in previous chapters.

It may be insisted that any kind of conceptual clarification or analysis which is not prepared to face revisionary elements is narrow-minded. Surely there might be good reasons to change our conceptual practices, and if so, why would we limit ourselves to clarifying concepts of only very limited usefulness? It might be illuminating to draw a parallel with illnesses and diseases here. As medical science advanced, we have started to use more sophisticated concepts for what was wrong with ill people. To diagnose somebody who is ill as 'possessed by demons' regardless of the precise symptoms does not admit of much more effective treatment than applying leeches to remove blood or getting a priest to exorcise the demon. In some cases this may have improved the sufferer's situation – or at least have seemed to correlate with such an improvement – but somebody who had contracted the plague would not generally have been cured in this way. A major advance in healthcare was made because of the insight that many illnesses are caused by micro-organisms, and that distinguishing between illnesses by complexes of symptoms by and large allows us to classify them according to the micro-organisms that they are caused by. The next step was to find out by what route such micro-organisms spread and infect people, in order to be able to stop epidemics. And yet further steps were made by researching how these micro-organisms attack the human body, finding ways to counteract the damage done and attacking or destroying the micro-organisms themselves. All these advancements rest on empirical discoveries, and enabled us to prevent, predict and control illnesses better.

There are three lessons that I want to draw from this example. Firstly, the advancements in preventing and curing diseases were made possible because we started to make use of natural kind concepts. The same cure does not work

against all illnesses, so the finer distinctions between them on the basis of which generalisations were applicable was a necessary step. Settling by trial and error which methods of cure and prevention work best is only possible if you have the conceptual means of distinguishing between the various cases in which they do and don't work. Secondly, there was a clear necessity, or at least demand, for such developments in conceptual practice, guided by empirical discovery: people were dying from these illnesses, so getting to grips with prevention and cure was very useful. And third, considerable revision of conceptual practices was necessary. To say of somebody that they are possessed by the devil is very different from saying that they suffer from (say) tuberculosis, even though they might be said of one and the same ill person.

How do these lessons translate to mental concepts? There is considerable disagreement over how useful or, on the contrary, how much in need of conceptual revision, folk psychology is. Paul Churchland[110] has argued that folk psychology and its conceptual framework leave so many things unexplained that eventually it will have to be superseded by a matured neuroscience. As science progresses, his argument is, we will come to appreciate that we are as mistaken now about the existence of beliefs, feelings and the like as we once were about the existence of phlogiston or witches. However, this argument assumes that our usage of mental concepts presupposes that they are natural kind concepts: erroneous conceptual practices are unmasked by empirical discoveries. (For example, phlogiston was unmasked as a concept without any reference because the assumption that combustion of any material was the same as losing the substance of phlogiston turned out to be tenable only if it had a negative mass, and no substances with negative mass exist.) Such a thing cannot happen if mental concepts are criterial concepts, if what I said earlier about the relation between natural kinds and natural kind concepts was right. In the parallel with illnesses above, science does not show that we were *mistaken*, but it suggests a more *useful* conceptual framework. Most philosophers do believe our framework of mental concepts to be very useful, but I know of no argument that could establish that such a development could not take place in (folk-)psychology. However, it could not be prompted by empirical evidence showing up any errors. Therefore, the absence of any obvious need to change our framework of mental concepts (which, if we take the parallel with illnesses seriously, would be likely to be quite radical) seems enough reason to stick to clarifying our current mental concepts. And about those, we have good reasons for resisting the causalist's conceptual claim.

---

[110] See, e.g., Churchland 1984

5.2.5 The essentialist claim

Will no claim inspired by the idea that mental kinds are natural kinds withstand scrutiny, then? Let us consider another claim, which in chapter 1 I have called causal essentialism:

> **(CE)** (Even though the concepts of psychological states may be innocent of causal implications, it may be that) mental states are essentially causally active or acted upon, or 'located in the causal swim'. The states and events we pick out by means of our mental concepts happen to have a causal essence of some kind, even though that is not reflected in those concepts.

There are two important aspects to this claim. Firstly, any claim about mental concepts is explicitly disavowed; it is a claim which is concerned only with the subject-matter of mental concepts. Secondly, a claim is made about some kind of essence of the subject-matter of mental concepts, namely that it is part of that essence that mental states or events have a place within a causal network.

Is it possible to make a claim about the subject-matter without making a claim about mental concepts themselves? This would be needed if **CE** is to be free of conceptual implications. To some extent what a concept refers to is a matter not only determined by the shape of the concept, but also by what the world is like. Take the concept of a car. It is no part of the concept 'car' that cars are red – cars can have all kind of colours -, and yet we can imagine living in a world in which all cars happen to be red. In that world, it would be true to say that cars are essentially red – anything that wasn't red, in such a world, would not be a car. Similarly it could be the case that, even though it is no part of the concept of perception that a perceptual state be caused by its object, all perceptual states happened to be caused by their object. It would be a feature of our world that it belongs to the essence of perceptual states that they are caused in that way – however, in another world perceptual states may be caused in another way, or perhaps not even be causal states at all.

It should be noted that not all claims of this type can co-exist with any conceptual claim. Suppose for example that it is part of the concept 'car' that cars have at least three wheels; if that is so, then it is impossible that the world would be such that all cars in it had two wheels. The two-wheeled objects in that world would simply not be cars but something else. However, the point I have been making against the **N(S)CA** claim is merely the negative one that mental states do not need to have a specific causal ancestry in order to fall under a concept such as

action or perception. So it could be that all my arguments were right and yet that the world be such that the subject matter of these concepts be caused in a certain way.

So far we have not made any use of the idea of natural kinds. The example I used for showing that something can be true of a subject-matter without it being reflected in the concept was that of red cars, presumably not a natural kind. But with the idea that the subject-matter about which we are making a claim forms a natural kind, the claim becomes stronger. For suppose that perceptual states form a natural kind by virtue of their causal ancestry. Then it would be the case that perceptual states necessarily - in any possible world - have such causal ancestry. Yet, if I was right about the possibility of using a criterial concept to refer to a natural kind, there is no incompatibility with causal ancestry not being a part of the concept of perception, even though it would be a bit of a cosmic accident if the reference of our criterial concept happened to coincide with a natural kind.

The claim that perceptual states (and, mutatis mutandis, the same would go for other types of mental state) necessarily have a specific causal ancestry looks like a strong claim. In order to appreciate how strong exactly, let me compare it with the conceptual claim that a state would have to have a specific causal ancestry in order to fall under the concept. Suppose that we assert that Pia perceives a platypus, but that it is demonstrated to us that there is no platypus causing a mental state in Pia. If the conceptual claim were true, we would then have to say that we were mistaken in using the concept of perception: whatever is going on, it is not something that can be called perception. By contrast, if the essentialist claim were true, it is not entirely clear what we should say. There would be some case for reconsidering whether what is going on should be called perception, just as there would be a case for reconsidering whether we should call a blue car a car if all hitherto encountered samples were red. But if there were good conceptual reasons for insisting that our case was really one in which Pia perceives the platypus – and if all I said about the conceptual claim is right then this could well happen – then we would have to admit that the world simply wasn't as we had supposed it to be, namely one in which all perceptual states have a specific causal ancestry. In other words, we would have to reject the essentialist claim after all. The conceptual claim always takes precedence. If empirical evidence plus conceptual claim together are incompatible with an essentialist claim, then the latter will have to go, which shows that the conceptual claim is a good deal stronger. Another way of putting this point would be to say that the essentialist claim is more of a hypothesis, which may inductively be supported by empirical

evidence but not proven true by deduction. The inductive support may however be very strong, which would make it likely that the essentialist claim is true.

The essentialist claim with which I started was weaker than the one I have discussed so far. It claimed not that mental states (essentially) have a specific causal ancestry, but only that they can be placed somewhere in the causal network of states and events. Obviously, a state may be located *somewhere* in the causal network – be in the 'causal swim' – without having the specific causal ancestry figuring in the other claim, and therefore this claim is weaker. On the – non-trivial – assumption that there is just one causal network, i.e., that causal pluralism is false, what it says is something to the effect that the subject-matter of mental concepts is not anything supernatural, or contrary to what physics is about, some immaterial substance. Thus it is a claim that folk-psychology – which uses the framework of mental concepts – can be true without committing us to another substance (Cartesian substance-dualism) nor to non-existing subject-matter (eliminativism). This claim, like the stronger one, is perfectly compatible with the conceptual claim that for a state to fall under a mental concept it does not need to have a specific causal ancestry. My claims to the effect that **N(S)CA** lacks support are evidently compatible with a monistic, materialistic philosophy of mind. But if **CE** is merely a statement to the effect that materialism in one shape or another is true, we may wonder if it deserves to be called a causalist position. It is, as I have shown, a good deal weaker than causal claims about the mental commonly found in the philosophical literature, and as such I see no reason to have any quarrel with it.

The argument from natural kinds is not the only causalist argument that can be understood as arguing for the weaker essentialist claim instead of the conceptual claim. The argument from reliability in the causal theory of perception (p.98), for example, could plausibly be understood as saying that reliable links in our actual world happen to be causal links. So if perceptual states reliably inform the subject about objects in the world, then such states are necessarily caused by those objects – assuming the world is as we assume it to be, namely, one in which reliable links are causal links. But, again, the stronger conceptual conclusion which Child attaches to this argument is not thereby available. However, the likely truth of the weaker essentialist claim may do much to explain why the conceptual claim has seemed so obviously true to many philosophers.

5.3 Causal distinctions

Besides **CE** there is another causalist claim which may well be plausible and explain some of the attraction of **NCA** claims. I did mention it in chapter 1 and have not discussed it further; it is the claim that there are certain distinctions to be made in the 'arena' of mental concepts which are irrecoverable if not made by causal means (**VCD**).

Suppose for example that one has a thought about an apple that is on the kitchen table. You have looked at it and weighed it, and express the thought: "The apple on the table is red and weighs 200 grams." However, the apple that was on the kitchen table has been eaten, and replaced by a numerically different apple of the same colour and weight. Now there seem to be two different possible thoughts that you might be expressing: it might be about the apple that has been eaten, or about whatever apple is on the kitchen table. They are obviously different, because the first thought is false whereas the second is true. How are we to distinguish between them? It seems natural to make the distinction in causal terms: you have had previous causal contact with the apple of which only a core is left in the bin, but not with the apple that is on the kitchen table now.

I do not intend to launch into a discussion here of the different kinds of thought that one may distinguish, and whether the causal condition is the right one for making such distinctions. Nor do I want to survey which other distinctions between mental concepts can only be made by means of a causal condition. The point that I want to make here is that I have not given any arguments against **VCD**. Nor is it the case that, should it turn out that there are distinctions that can only be made in causal terms, this automatically provides a conclusive argument in favour of **NCA** about perception, or **FPCE** about action. My limited conclusion have been that there is no conclusive argument in favour of some special cases of **VCD**: namely, that the distinction between hallucination and genuine perception, between actions and mere movements, or between the real reason and the mere justification of an intentional action, need to be made in causal terms. This leaves open that there are other distinctions legitimately to be made in causal terms, although the conclusions that I have drawn are good reason for cautious scepticism. But a full survey of such other distinctions, and their relationships with the ones I argue against in this thesis, is a follow-up project.

5.4 Summary and conclusions

In this chapter I have discussed two important causalist arguments which had not been decisively rebutted in earlier chapters.

The argument from counterfactuals argues that there must be a causal link between an action and its primary reason, or a perceptual state and its object, on the grounds that in the genuine cases a counterfactual is true. That counterfactual is made true by a causal link, so it is argued by the causalist. I replied, firstly, that there is an obvious disanalogy with acceptedly causal counterfactuals. 'Causal counterfactual', in this context, must be understood as one *made true by* a causal link between antecedent and consequent; it is not enough that there be such a causal link. Secondly, I have shown action explanation to make true other counterfactuals than the one which might be interpreted to be causal. Accepted causal explanations do not make true those kinds of counterfactuals which, since they have to do with the goal-directed character of intentional action, seem to be central to action explanation.

The argument from natural kinds is based on the idea that mental kinds may be (discovered to be) natural kinds. Mental kinds may indeed be natural kinds, although not – if we accept the multiple realisation thesis – of the type that in effect reduces the mental to the physical. However, we cannot argue from this to the conclusion that mental concepts are natural kind concepts. Such a conclusion needs an argument of its own, which is not based on empirical discoveries: but that brings us back to the discussions of previous chapters. The argument from natural kinds does therefore not bring in any new considerations regarding conceptual causalist claims. It does however suggest a weaker claim, which I have called causal essentialism (**CE**). This claim, which asserts that the subject-matter of mental concepts consists of states and events that are essentially causally active and acted upon, has turned out to be defensible.

6 Diagnosis and conclusions

In this chapter I want to tie the ends together. Firstly, I want to briefly review how the different causalist claims that I have distinguished in chapter 1 have withstood the various arguments that were marshalled against them in chapters 2, 3, 4, and 5. This then clears the way for a tentative diagnosis of what is at the root of the types of causalism that I have been attacking; I will suggest that a particular interpretation of the adagium that 'the mind represents the world' (in one word: representationalism) is at fault. I discuss reasons to endorse representationalism, and reasons to reject it. Then I attempt an assessment of the predicament of current philosophy of mind, if one is to accept my arguments against causalism and representationalism. I try to sketch what a future philosophy of mind may look like.

6.1 Where does causalism stand?

I started out, in chapter 1, by distinguishing a number of causal claims. How have they fared throughout the argument in this thesis?

I have not quibbled with any results of empirical science. So the claim that causal processes are involved when we act, perceive, etc., (**CPI**) is uncontroversial. The stronger claim that the subject matter of mental concepts essentially consists of causally active states (**CE**) has not been contested either. I have not bothered much with the claim that there are certain valid distinctions between mental concepts that can only be made in causal terms. However, I have strongly resisted the claim that causal ancestry or progeny is a necessary condition for application of the concepts of perception and action (**NCA**, **NCP**). On that matter, the ball is in the causalists' court. I have also argued against the claim that folk-psychological explanation must be causal explanation (**FPCE**).

6.2 Representationalism and causalism

As I argued in 4.3, the objections I directed at causal theories are, curiously enough, not dependent on any specific account of causality. Is there an idea deeper under the surface driving the positions I object to? I think that that causal claims such as **N(S)CA**, **FPCE** and **FPCS** are motivated by what we may call representationalism. The term 'representationalism' I will here use to refer to the thesis that

> (**REP**) For someone to have a mind they must have inner mental states which are representations of states of affairs in the outside world.

What sets the mental apart from the physical, the thought is here, is that it has *aboutness* or *intentionality*. (At a pinch this could be said of Descartes as well, who emphasises that thinking is the essence of the mental substance. For thinking is concerned with the ideas which the mind receives of the material world.) How else than by representing that which is outside[111] the mind can minded beings engage in intelligent behaviour that is adaptive to their environment? Minds and minded beings distinguish themselves from the physical in that their actions are not straightforwardly determined by outer occurrences and influences. They exhibit not mere physical reactions but engage with the world in an adaptive manner. In order to be able to do this they must somehow be able to represent the current state of affairs in the world, but also possible states of affairs.

The mind, then, represents the world. But far from this being the point where the story ends, this is where it really begins. For we haven't said much yet, in attributing to the mind a capacity to represent: the philosophical challenge is to achieve some understanding as to how it can come to do such a thing. The general question, 'how can one thing represent another?', is not easily answered.

In the mental case the usual answer is that there are ideas or mental states which both resemble, and are caused by, what they represent. They must resemble: a painting resembles what it represents, and it is hard to see how a mental state would be about, say, a cat, if there were nothing 'catty' about it. It is of course unclear how literally we need to interpret this requirement of resemblance – we need not assume that there is a furry purring animal in my mind. Indeed, some representations don't seem to resemble at all what they represent, at least not in a literal sense: the word 'cat' is not a furry 4-legged animal. Apart from this, how the resemblance has come to be needs to be explained, and this is what the causal requirement does. The imprint of a rubberstamp resembles the stamp (the stamping surface of it, at least) by being caused by it. Why this should be a causal requirement is a question to which I will return below. The causal requirement by itself is not enough, because often enough causes and effects don't resemble each other in the slightest, nor do effects always represent causes. My flat bicycle-tyre may have been caused by the broken glass on the road, but it does not represent it.

What, more precisely, is the relationship between the representational theory and the causalism that I have been attacking? The **NCA**-thesis about perception is

[111] There is in **REP** an implicit commitment to what McCulloch calls the thesis of self-containedness of mind with respect to world: "[minds] are capable of having the mental characteristics that they do, independently of the existence of any body." (McCulloch 1995 p.11) This commitment may not in its explicit form be acceptable to all modern causalists.

just a part of the representational theory. For, firstly, it fleshes out what needs to be the case for a state to be a mental state (of a specific kind). Secondly, it does so in terms of a causal condition, which we saw to be one of the conditions for one thing representing another. (To deny **NCA** but uphold **REP** one would have to deny that the causal ancestry condition is necessary for representation. That combines badly with the thesis of self-containedness.) Other causalist theses are less directly related, but do start from the assumption that mental states are the sort of things that can cause and be caused, and that they are representational states. Beliefs and desires represent the world in their different ways, and have to be causal items to figure in causal reason-explanations (**FPCE+FPCS** about action). **NCP** about action has a hidden commitment to representationalism: although the condition is in terms of causal progeny (a trying is an action if it causes a bodily movement of the right sort), there is an appropriateness condition. It is not sufficient if the trying causes no-matter-what, and to be able to give hand and feet to this the trying must represent (a) state(s) of affairs. **NCA** is some kind of mirror-image of this, since it is the movement and not the trying or volition which is an action if the causal condition is fulfilled, but the volition needs to be a representational causal item for much the same reason. **FPCS** is entirely explicit in its requirement that mental states are representational causal states. The only exception seems to be **CPI**, which says nothing about the representational character of mental states, nor says anything about what must be the case in order to be minded or have mental states.

Notice, that there are really two components to representationalism. The first is that mental states are inner representational states; the second, that this representational character is due to their being caused in a certain way. Once the first is endorsed, it is hard to resist the second. However, the thesis that our ability to represent our environment is essential to our having a mental life arguably does not force us to conceive of the way in which this representing takes place as by means of inner representational states. The step is obviously a natural one, and easily made, as demonstrated by Tim Crane, when in the introduction to his "The Mechanical Mind" he writes about

> "...the philosophical problem of mental representation. This problem is easily stated: how can the mind represent anything? My belief, for example, that Nixon visited China is about Nixon and China - but how can a state of my mind be 'about' Nixon or China? What is it for a mind to represent anything at all? For that matter, what is it for anything (whether a mind or not) to represent anything else?" (Crane 1995b, p.1)

It starts off innocently enough, asking 'how the mind can represent anything' (although to ask, 'how do we represent things' would have been more neutral),

but then suddenly the question becomes: how can a state of my mind be 'about' something?

In the following sections, the following question will be central: is there a way to allow for the intuition that representation is important, perhaps even essential, to mentality, without committing ourselves to the 'inner states'-picture that drives causalism? Before that, however, I will go deeper into the question why the causal link should be a requirement for representation, given the inner-state conception. This will turn out to be a motivation from science, but not altogether a recent one.

6.2.1 Representationalism and metaphysics

Above I have indicated that the causal requirement, i.e. the requirement that mental representations be caused by what they represent, was needed to explain how the representation came to resemble what it represents. The causal condition, however, is not the only possible way to explain how two things can resemble each other. Another explanation might run along lines made popular (among philosophers) by Plato. The idea would be that were one thing resembles another, they both partake of the same Form. Two cats resemble each other, because they both partake in the form of Cat-hood; of course they do so imperfectly, so they are not entirely the same. A cat-representation in my mind also takes partakes in this Form, in a different way, and that explains the resemblance: for the same reason that two cats resemble each other, my mental representation may resemble a cat.

There are at least two problems, though, with this kind of explanation. Firstly, it is not particularly *illuminating*. To say that two things partake of the same form seems just to be a roundabout way of saying that they resemble each other, and so does not explain anything. However, if it does say something else, then more needs to be said about what exactly that amounts to. Secondly, it does not seem very *scientific*. In Descartes' day, the idea – also due to his contemporary, Galilei – became dominant that the laws of mechanics would serve to explain scientifically everything worth explaining. For Descartes and many philosophers after him this included minds. An explanation in terms of Forms does not fit such a mechanistic pattern; the need was thus felt to give a mechanistic explanation of representation and the involved resemblance.

However, Descartes faced a problem: on the one hand, minds had to be made part of the scientific world-order, and therefore mechanistically explained. On the other hand, minds cannot just be like machines or clockwork, with perhaps only

their complexity as a differentiating factor: this conflicted with his religious and moral convictions. He adopted what Gilbert Ryle[112] has called the para-mechanical hypothesis: minds are at once very similar to machines (because explained mechanistically), but at the same time very different because of a different substance. "Minds are not bits of clockwork, they are just bits of not-clockwork."[113]

The philosophy of mind has of course developed since the 17th century. Where Descartes thought that the mind was a substance on its own, thereby embracing substance-dualism, it is currently fashionable to be of monistic[114], materialistic outlook. However, the predilection for mechanistic explanation has stayed. The currently dominant project of physicalism is to show that mental states are somehow related to physical states; and thus, that there is nothing that falls outside of the scope of (mechanistic) physical explanation (but the idea that mental explanation reduces to physical explanation is nonetheless resisted). With this mechanistic outlook comes the conception of the mind as a representational engine, operating by means of internal causal states, which is what I have been arguing against. For now the point to note is that this conception is motivated by a desire to fit the study of mind with developments in contemporary science, and that this motivation basically has not changed since the 17th century, even if much else in the philosophy of mind has.

6.2.2 Representationalism, for and against

In the following sections I want to see if we can escape some of the attraction of this representationalist picture, and to see if we can account for some of the uses we make of the notion of representation without referring to an inner representational state.

6.2.2.1 Goal-directed behaviour

Perhaps the most vivid argument in favour of representationalism is this: intentional agents exhibit purposive, goal-directed behaviour. What prospects could there be for explaining that kind of behaviour if not by reference to what agents want and believe? And are such beliefs and desires, in order to make that kind of explanation possible, to be conceived of as inner representational states?

---

[112] Ryle 1949

[113] Ryle 1949, p.21

[114] A monism of substance that is, not necessarily of properties, as I have argued in the section on mental causation in chapter 2.

Collins says the following about reason explanation:

> "In the interpretation of reason-giving put forward here I press for the elimination of any role for the fact (where it is a fact) that the agent wanted to attain the objective, reference to which explains his action. Of course, I do not deny that agents commonly do want to reach the objectives that their actions do reach." (Collins 1984, p.144)

The basic question to pose to the representationalist is this: why should it be supposed that beliefs and desires, conceived of as internal states, are explanatory of behaviour? Just as easily, the picture could be inverted: that people exhibit goal-directed behaviour is a given fact, something that does not need explaining. And based on the goal-directed behaviour that they exhibit, we can attribute beliefs and desires to people. Action explanations, on that view, may *appear* to explain a particular action in terms of internal representational beliefs whereas what they in fact do is to explain it by describing it as an instance of a certain kind of goal-directed behaviour. Describing an agent's action as such an instance would then be done by ascribing a certain belief-desire combination to him.
In the next section, hopefully it will become clearer what the alternative picture proposes.

6.2.2.2 The possibility of error

A powerful reason for thinking that there needs to be an internal representation in us is that we can be mistaken about what the world is like facts about the internal life of the agent rather than facts about the world are explanatory.
In the case of intentional action, it is tempting to reason in the following way. The agent represents the world to himself in a certain way. But of course that representation can be mistaken. When this occurs, there is no 'external' state of affairs that can explain the action; therefore, an internal representation must (depending on the particular flavour of causalism) cause or explain the action.
For example, let's say that I apply a special wax to my car in order to protect it from rusting. Assume that in fact this wax has been sold to me by a clever salesperson, and that it doesn't protect the car at all. Still one might argue that I apply the wax to my car because I *want* to protect it from rusting, and I *believe* that applying this wax will do so. Whatever facts I am ignorant about - the effectiveness of the wax, or why and how cars rust - this explanation will stand no matter what. The explanation in terms of beliefs – the argument goes - must be the more basic one, because the explanation in terms of facts holds only in special cases, namely those in which the belief is true and matches the facts. Additional support for this stance comes from the observation that, in case the

wax *does* protect my car from rusting, but I *don't* believe that it does, it doesn't seem right to explain my action of applying the wax by saying that I do so in order to protect it from rust - even if that's what it does.

So the internalist argument is basically as follows: Whether or not the wax is protecting my car against rusting, I am performing the same action: namely, applying the wax to my car. It seems logical that there must be a *common factor*[115] in the explanation of these two cases: namely, my belief that this wax protects my car. The thought is that the belief, veridical or non-veridical, explains why the two cases feature the same action; the relevant facts cannot do this since they are different. However, this is putting things the wrong way around: mentioning the agent's belief is a way of explaining *in what respect* the two actions are similar, not *why* they are similar. What I am suggesting is a very different picture. I do not, of course, deny that agents have beliefs, nor that they act accordingly; but beliefs are not explanatory items. Actions, on this picture, are explained by reference to what they are directed at. Actions are criteria for beliefs: they justify us in attributing certain beliefs to the agent. On the basis of what an agent says and does, we form a hypothesis about his beliefs.

Looking at the matter in this way, we can understand what is going on in the case which seemed to lend additional support to the internalist stance, namely the one where the wax does protect my car from rusting and I apply it to my car, even though I do not believe that it so protects my car. The action prima facie justifies attributing to me the belief that the wax protects the car. However, suppose someone were to tell me that in fact it doesn't do so, and that I would nonetheless cheerfully go on with the job. Moreover, if I were to find that my car started rusting soon afterwards, I would not go back to the shop to complain about the product, nor would I try out some other product instead.

These counterfactuals –the importance of which I pointed out in section 5.1.1 - do two things. Firstly, they defeat the belief attribution. Secondly, by the same token, the explanation that was given of my action ('in order to protect my car from rusting') is falsified, because the truth of that explanation implies the truth of certain counterfactuals, conflicting with the ones that are actually true. What might be the case instead is that I apply the wax in order to create the impression that I am a busy man, or looking well after my car: such explanations, if true, imply a range of other counterfactuals, and matching beliefs can be justifiably attributed to the agent.

[115] This terminology is not a coincidence: the points made here for & against 'going internal' are the same as in the discussions about causal and disjunctive theories of perception.

What about the other case, that in which the agent acts according to a false belief? Let us reconsider the envisaged explanation of this action: namely, that I apply the wax in order to protect my car from rusting. There is no obvious reason why it should become false because of a fact about the wax. The essence of goal-directed behaviour is the compensatory character of it, not that it actually reaches its goal. If it is the holy grail that I'm looking for, that doesn't make my quest devoid of any purpose.

Now it may be that I have given some plausibility to the claim that we can explain actions without reference to an internal representational state, even in the case where somebody acts on a false belief. But even so the intuition that we can attribute false beliefs to somebody – whether or not that is needed in order to explain their actions – is still standing. How is that kind of talk, on a non-representational outlook, to be reconstructed?

Let us look at the grammar of 'representing'. I contend that when something - an object or a state of affairs - is represented, it is always represented to something or somebody, which or who can understand or interpret the representation. A personal gift may represent a particular occasion or person to me; an imposing statue may represent dictatorship to citizens; a professor may represent perfect scholarship to his students. Often, however, it is not stated who is represented to, because it represents to whoever cares to take notice of and interpret it in accordance with conventions. A picture can represent a landscape (to whoever looks at it); ink marks on paper represent letters (to whoever reads it); a scale reading represents a certain weight (to the grocer and his customer).

Now consider the following examples: "He went around the field because he thought there was a bull in it" (but in fact there was not), or: "He stretched out his hand because he seemed to see a cat, and wanted to stroke it" (but none was there). Now what could we mean by saying "..he thought there was a bull in the field"? A philosopher might say, "he represented the world as being such that the field contained a bull." Who did he represent this to, and how? The standard interpretation has it that in the man's mind there is a mental picture of the field with a bull in it (or perhaps a picture of himself, having gone into the field, and being on the run with a bull on his heels. In fact, it seems odd to think that the representation involved is a private mental one, only accessible to the agent himself. A non-representationalist might construe the matter thus: the man's doing what he did (namely walking around the field rather than across it) represented the world as being such that the field had a bull in it, to an onlooker (somebody who might want to explain his action, maybe) or to himself. The man who stretches out his hand to stroke a non-existing cat represents, in doing so, the

world as having a cat in it near him; he thus represents it to anyone who sees him do this, and/or to himself. In other words, it is the action, not some internal state, which is the representation.

One may object in the following way: how could we know that the man's action represents the field as having a bull in it, rather than an angry farmer who is likely to deny him right of way, or as being too soggy to walk across it? Well, are we ever indefeasible in judging what something represents? Presumably not; there is often more than one way to help figuring it out, and even so, one may be just make a plain mistake. But we very rarely attribute a thought explaining an action on the basis of just one observation - more context helps. Did he ever before have trouble with an aggressive bull? Is he a fanatical right-of-way campaigner? Is he acquainted with the land-owner? Does he try to stop his companion crossing the same field? Has he crossed this field before? Does the ground look wet? What kind of shoes is he wearing? Does he wear anything red? Does he cross the next field, which has a sign at the fence, "Beware of the bull"? If he is asked why he went round the field instead of crossing it, what does he answer?

In the normal course of explaining someone's action we almost effortlessly take many facts of this kind into account; so much so, that we are tempted to assuming that our assessment must be far less sophisticated than that of the agent himself, who (on the traditional view) after all just has to introspect to find the answer to the question, "why did I do that?" To change tack: why should we assume that there is always a definite answer to this question? I have, at any time, many beliefs; it may not be very clear to myself upon which beliefs I acted when I was making my choice, so why should it be transparent to someone else? Is there even a definite fact about the matter? When one asks somebody for an explanation, usually some answer will come out: but may this not always be what Davidson calls 'mere rationalisation'? It could well be the case that we attach more importance to rationality than it has - and that in doing so we are conditioned to expect, and give, rational explanations of actions.

6.2.2.3 Cognitive science

It is often insisted that to posit representations is of crucial importance to the advancement of cognitive science (see e.g. Gardner 1985, Sterelny 1990). Does my view that we'd better conceive of mental representation other than explained by inner causal states go against this, thereby disregarding what fruitful work there is and has been in cognitive science? Gardner writes: "Whoever wishes to

banish the representational level from scientific discourse would be compelled to explain language, problem solving, classification, and the like strictly in terms of neurological and cultural analysis. The discoveries of the last thirty years make such an alternative most unpalatable."[116]

The almost universally recognised importance of representations originates, perhaps, in Marr's study of perception[117]. He posited three levels of explanation: computational theory, algorithm, and implementation. Sterelny, whose terminology I prefer, calls these levels respectively the ecological level, the computational level, and the level of physical implementation. The most abstract level specifies *what* the cognitive capacity being studied does in functional terms, and *why* (in terms of survival value, e.g.), but without prejudging the issue of *how* it does that. The intermediate level explains in terms of representations being processed according a certain algorithm how this function is realised. At the level of implementation, finally, the neurological or physical realisation of these representations and algorithms is explained.

So far, so good: it seems sensible to posit an intermediate level of abstraction if we want to study the workings of a cognitive capacity, in the hope to eventually simulate or duplicate it. But notice that we have talked here only of a *level of explanation*, i.e., a tool used for studying a phenomenon. There is no committal to any ontological status for such a representation. If I want to study the path of a river flowing downhill, I can make a similar move. At the ecological level, we say that the river transports water to the sea. At the representational level, a story can be given about the profile of the landscape between the source and the sea, and how a representation of it gets processed in order to find downhill gradients. Finally, the implementation is based on following the path of least resistance as shaped by gravitational force. Sterelny makes a similar point about a mechanical clock: "We do not regard the clock as executing an algorithm that calculates the time, and then try to explain the workings of the devices that implement such executions."[118] The conclusion is that in cases like these, the computational description is "surely otiose". However, how can we be so sure that the case is different for cognitive/mental phenomena?

> "Why then cannot we rest content with two levels: an ecological level telling us what the mind can do, and a neurological theory, or suite of theories, giving an account of mechanism. What is the point of a *computational* level? [..] I have no entirely convincing solution to offer. But, very roughly, a computational explanation is required when the system to be explained can recognise the same information in a wide variety of physically distinct

[116] Gardner 1985, p. 383
[117] Marr 1982
[118] Sterelny 1990, p.49

> signals, and/or use that information for a range of purposes. That is, we explain a system computationally if it shows the informational sensitivity [which is] symptomatic of sentience." (Sterelny 1990, p.50)

Sentient or intelligent behaviour can adapt itself to new circumstances, and it is sensitive to the informational content of sensory stimuli, not to their physical format. But again, this only gives us a principled reason why the use of a certain tool - description at a computational level - is appropriate. Unfortunately, to one specific function there correspond several possible different computational explanations, and to that in turn different possible implementations. Simply by studying the cognitive phenomena, then, we cannot tell which is the 'right' computational explanation, even though certain hypotheses will be excluded by empirical evidence. There may not even be an answer to the question: "which is the right explanation?"

It should be obvious which conclusion I would like to draw from these considerations. Whereas representations do fulfil a useful *explanatory role* in cognitive science, there is no reason at all to think that in order to do so we must conceive of them as inner causal items. The alleged successes and empirical evidence of cognitive science are not forcing representations on us in the philosophically pernicious sense: the use of representations as mere explanatory tool in cognitive science is compatible with rejecting the philosophical thesis that representational items of a certain character must be causally involved in certain phenomena for them to be of the right mental kind.

6.2.2.4 Scepticism

On this point I can be short, since to do otherwise would be to repeat what I have said elsewhere. **REP** leads to scepticism of at least the following two kinds. If the visual experience involved in perception is a representational state caused by states of affairs in 'the outside world', then we can never be sure that we have any knowledge about that world, or even be sure that it exists. And again, if the bodily movements involved in actions are caused by representational inner states, then we can never be sure that they are anything more than bodily movements, or even that other agents exist at all. These kinds of scepticism, if they arise because of **REP**, surely are a reason to reject that thesis.

6.2.2.5 Externalism and the efficacy of content

An important debate in the philosophy of mind that has been raging ever since Putnam's "The meaning of 'meaning'"[119] was published has also bearing on the discussion of representationalism here. Putnam argues, by means of his famous "Twin-Earth" thought experiment, that 'meanings just ain't in the head': the content of mental states is to be individuated at least in part by reference to states of affairs outside the subject. I will not reproduce his arguments for that thesis here, but simply acknowledge that many philosophers have been convinced by this line of argument, and have become 'externalists'.

Now the following problem arises for the causal representationalist: if representational content is supposed to play a causal role in a subject's mental phenomena, but that content is not individuated solely on the basis of states of affairs internal to the subject, and therefore not completely located within the subject, how can it play that role? Surely, a state of affairs not directly influencing the physical basis of mental representations cannot influence their causal role?

Now one may well try to resist this argument. Martha Klein[120], in an excellent discussion, points out that it has not been established that the externalist would be committed to holding that a difference in representational content must imply a difference in causal powers. Be that as it may, the debate has become one about whether we should be methodological solipsists or not. Methodological solipsism is a term coined by Putnam:

> "When traditional philosophers talked about psychological states (or 'mental' states), they made an assumption which we may call the assumption of methodological solipsism. This assumption is the assumption that no psychological state, properly so called, presupposes the existence of any individual other than the subject to whom that state is ascribed." (Putnam 1975, p.220)

Putnam proposed to call mental or psychological states that are permitted by methodological solipsism mental states 'in the narrow sense', and those not permitted by it, mental states 'in the wide sense'.

The Twin-Earth thought experiment is widely taken to have demonstrated that if we want full-blown representational content to play a causal role in the mental, then mental states in the wide sense must be admitted into the causal order. So it seems that in order for mental content to play a causal role, methodological solipsism must be false.

---

[119] Putnam 1975
[120] Klein 1996

Michael Devitt[121], among others, argues that one can both accommodate the externalist thesis that mental states are individuated at least in part by reference to states of affairs outside the subject's skin, *and* retain methodological solipsism. The 'trick' consists in distinguishing between two kinds of content that a mental state can be said to have: 'narrow' content and 'wide' content. Wide content is the truth-functional kind, the kind that has meaning in the usual sense. Narrow content is a dressed-down notion, stripped of its truth-functionality, but it is all we need for the causal role of the mental state. (Notice the seductive similarity with Putnam's use of the terms 'narrow' and 'wide'. The point made in the two cases is, however, completely different. Whereas for Putnam something is either a mental state in the wide sense or a mental state in the narrow sense, Fodor, Devitt, and others hold that any given mental state can be said to have two different kinds of content at the same time: narrow and wide.) The point about causal role seems correct: if I have exactly the same brainstate as my twin on Twin Earth, the same bodily movements (etc.) will ensue, even though my thought was "I would like a drink of water" and my behaviour constituted drinking a glass of water, whereas my Twin thought "I would like a drink of twater" and engaged in twater-drinking behaviour. However, why think of narrow content as content at all? Nobody ever made the claim that the state playing the causal role was not within the subject; therefore, if a mental state's narrow content is taken to be the same as its correlated brainstate's playing a certain causal role, there is no real dispute. But the claim that truth-functional, 'wide' content – the kind of content that mental states have - is not in the head, still stands. The externalists were talking about mental phenomena, but internalists have changed the subject matter to brainstates. The situation might be clearer if externalists insisted that to the extent that meaning is not in the head, the mind is not a container of representational states.

6.2.2.6 Mental causation

We have seen (2.2.5) that mental causation is a big problem in contemporary philosophy of mind. However, if we reject **REP** with its conception of mental states as inner causal items that represent the outside world, then much of the need for mental causation disappears as well.

The problem of mental causation is that of giving an account of how the mental can be part of the causal order *qua* mental. To refresh your memory, I again

[121] See Devitt 1990 in Lycan 1990

quote Crane, for a statement of the five theses that are inconsistent when taken together:

> "(A) Causes have their effects in virtue of (some of) their properties
> (B) There is mental causation
> (C) The completeness of physics is true
> (D) There is no overdetermination
> (E) Mental and physical causation are 'homogeneous'."
> (Crane 1995a, p.229)

This formulation leaves out a vital metaphysical assumption, namely that the mental is realised in, but cannot be reduced to, the physical. Now, is anybody who subscribes to some form or other of non-reductive materialism - and I count myself among them - stuck with an insoluble problem?

My proposal, then, for how to (dis)solve the problem of mental causation is that we should reject the thesis that there is mental causation. Crane writes about such a position: "An epiphenomenalist will reject (B) [the thesis that there is mental causation]. But surely this is the last assumption we should reject - that our minds make our bodies move is not a piece of philosophical theory, but something which theory should explain." (Crane 1995a, p.230) In opposition, I want to make plausible that we can reject mental causation, without flying in the face of common sense. To persist in branding such a position as epiphenomenalist, a position according to which "the mind doesn't matter", is at best misleading.

It is vital that we consider what exactly it is that common-sense says in order to pick apart what on the one hand needs to be explained by philosophical theory and what on the other hand falls already firmly within (misguided) philosophical theory. Let us distinguish three kinds of statements that can be made about people and their behaviour; I shall indicate them by example.

(1) "The heat of the stove made him withdraw his hand", or "Sentimental films always make me cry", or "The stock market crash caused investors to sell in a panic". (We could call statements of this kind causal-behavioural.)

(2) "Tissue-damage in general causes C-fibres to fire in humans", or "Retinal stimulation causes activation of the visual cortex" (causal-physiological).

(3) "He borrowed a ladder because he wanted to take a look at his leaking roof", or "I took a taxi because I wanted to catch a train" (reason explanation).

All three kinds of statements can, I think, legitimately be made; common-sense - or, in the case of (2), science - has a straightforward use for them. That is to say, none of these are philosophical jargon which we might have to be wary of. Each of them, though very different from each other, may be thought to support the claim that "there is mental causation". However, none of the three does so, at

least not in the sense that philosophers have been concerned with. The third kind of statement, reason explanation, is widely thought to be causal explanation because of its form, but as we have seen a closer scrutiny reveals that it is teleological instead of causal. The second kind of statement is causal, but no mention is made of mental states, events or processes. It is, indeed, the availability of this kind of statements which is thought to threaten mental causation, because they appear to point out that all the causal work is already done at the physiological level. (Continuing this line of thought, physiological causation is then in turn under threat from physical causation, and so on.)

But what of the causal-behavioural kind of statement? Is there no mental causation there? They appear indeed to be causal statements, and in the sense that they are about people's behaviour they seem to merit the label 'mental'. However, this is not the mental causation that worries philosophers. The 'mental causation' which is problematic is between mental states - beliefs, desires, emotions, sensations and so on - and the physical world - hot stoves, moving limbs, etc. "That our minds make our bodies move is not a piece of philosophical theory", according to Crane. However, I think that it quite obviously is. It is born out of the thought that for causal-behavioural statements to be causally acceptable, there must be an identifiable chain of efficient causation - a philosophical move if ever there was one. That leads to talk of causally related mental states, embodying the causalist/functionalist picture which I have been opposing. For that efficient causal chain to be acceptable, there must, in turn, be an underlying level of efficient causation. And then we are in trouble, because each lower level annihilates the work done at a higher level.

The usual approach to the problem of mental causation is to find some fault with the last step in the argument that I have just sketched. Philosophers have been quite desperate to show that mental causation and physical causation can co-exist quite peacefully. But a true solution must lead us away from the trouble before it starts: hence, my insistence that there is no need for mental causation (in the philosophers' sense) in the first place, and therefore we do not run into trouble having to account for its possibility.

"But surely the mental must make a difference; epiphenomenalism is an absurd position!", it will be objected. Of course: the mental *does* make a difference. Minded and unminded beings behave in very different ways; reference to people's mental lives enables us to explain, make sense of, and understand those ways. That makes a big difference indeed, but it forces nobody to think that there is mental causation, in the sense that inner representational states cause bodies to move.

6.2.3 Representationalism: conclusions

I have argued that representationalism is not needed (goal-directed behaviour; the possibility of error), and also that it leads to problems (scepticism; externalism; mental causation). Therefore it seems that there are no very convincing reasons for maintaining it as a driving factor behind the causal claims that I have been concerned with.

6.3 Non-causal philosophy of mind

Let us suppose, for the sake of argument, that causal claims such as **NCA** and **FPCE** turn out to be false. (Even though I have not established this claim, I have hopefully given it some plausibility.) What, it may be asked, have I got to offer as alternatives? The following is a brief speculation.

6.3.1 Mental concepts

The question driving causal claims about the conditions of application of mental concepts was: "how can we distinguish between an action and a mere bodily movement, or between genuine perception and hallucination?" I have argued that the argumentation for causal conditions is not conclusive. One might argue: if it is not a causal condition of some kind that answers our question, then what other kind of condition? I think that this move should be resisted. The way to do this would be by resisting what I called the driving question, and my arguments can be read as giving several reasons not to ask it.

In the first place, it is not obvious that the two are likely to be confused. To perceive something as an action is to see it as placed within a certain context, whereas to see it as a mere bodily movement is to abstract from that context. Secondly, if we interpret the possibility of confusion as entailing that there is a common element between the two, then we are stepping into the quicksand of scepticism, to which the causal theory is the doomed attempt to pull ourselves out of by the hair. Thirdly, the possibility of mistaking one thing for another does not in general require a common factor, and nor does a common factor automatically explain such a possibility.

In the case of action the alleged problem is to distinguish between two kinds of thing, actions and mere events. But seen in this light the question is not commonsensical at all; one is reminded of riddles such as: "What is the difference between a banana and a rabbit?". To put the point in a less rhetorical

way: it is like asking, "what is the difference between money and pieces of paper?" The concept of money is understood from within the practice of economics and trade, as e.g. playing a certain role in transactions. But it can be implemented by any kind of paper, metal, or whatever; this fixation on a physical basis would just distract us from the interesting questions that can be asked about money. I want to suggest that the same is the case with action. With perception again the dogmatic answer, on the assumption that there is a common element, is in terms of a natural relation, which has in fact no place among mental phenomena. (Compare: 'the difference between money and goods is that there is some kind of causal chain between money and the central bank.' Would anyone be tempted to argue for such an account?)

The suggestion that I would like to make is that the way to give an account of mental concepts is to flesh out exactly in what way the different mental concepts are interrelated. What needs doing is to explore how the concepts of action, perception, belief, knowledge etc. hang together; to explore what the differences and similarities between them are. This corresponds to how we learn to use the concepts of (to stick to our example) economics: if someone does not understand the concept of trade deficit, it is of no use to relate it to non-economical concepts - instead, what we need to do is to gain a better understanding of economics in order to be able to point out the place of a concept on the 'conceptual map'. What is needed, then, is careful conceptual cartography within folk psychology. It is quite amazing how neglected this task of conceptual cartography has been, because of the obsession with trying to make something like causal-ancestry accounts work for the central concepts.

In other words, I want to suggest that the move away from causal ancestry theories has to start from the recognition that what is to come in its place is a different type of account, concerned with answering different questions. To stay with the old questions will make it impossible to escape from the attraction of causalism. In part, the conceptual cartography will involve dividing up some concepts into finer-grained ones. It may be that the traditional division of the mental in action, perception, belief, knowledge, memory and sensation is rather arbitrary, and the result of overly broad brush-strokes. Certain mental states, for example, seem to hover between perception and belief: what one sees may be affected by what one believes. Knowledge and memory are terms that are to some extent interchangeable. Sensation is a concept capturing a very broad range from bodily sensations to emotions. Action, too, is a very broad concept, capturing a range from simple direct actions to complex communicative actions.

Now, the network of similarities and differences between sub-concepts arising out of this conceptual cartography may seem to blur rather than clarify the picture. The more detailed the map, the more complicated it becomes: but that should not surprise nor abhor us. The neat picture, in which similar causal ancestry claims could be given for the different mental concepts sitting side by side, has vanished anyway: for that was an artefact of the causalist view.

The exercise in conceptual geography would be something superficially not unlike what functionalists have been doing. Indeed, much of what functionalists say about the conceptual organisation of the mental is very commonsensical and hard to deny. But functionalism does involve a further step, which should be resisted:

> "A Wittgensteinian could even accept the broad outlines of Functionalism. Perhaps one can characterise believing in terms of inputs, outputs and other consequences. But this leaves it open to construe 'inputs' and 'outputs' and everything else in form-of-life terms (...) The thought that mental characteristics need to be located in a broader framework in itself leaves all the hard decisions about how to describe the elements of that framework still to be made. For the same reason, Wittgenstein does not have to deny that folk psychological descriptions help explain and predict behaviour. (...) One aspect of Functionalism of which a Wittgensteinian has to be more wary is the claim that beliefs and desires are causal entities in their own right whose nature is exhausted by their causal liaisons." (McCulloch 1995, p.121)

Functionalists think that mental states are to be placed in a *causal* network, and that the place of a state in that network decides what kind of state it is. For example, pain is (typically) caused by bodily damage, and typically causes damage-avoiding behaviour, or wound-tending behaviour. But notice two differences with my position. Firstly, functionalists are in the business of relating (mental) *states*, not *concepts*. Secondly, and relatedly, the relations that the functionalist is looking for are causal, whereas we should be looking at relations such as dependence, implication, and necessity. One way of bringing out the difference, perhaps, is to say that the project of the conceptual geographer is semantical rather than metaphysical. Of course, the functionalist's idea is that the semantical mirrors the metaphysical: once the metaphysical relations between mental states and processes have been sorted out, the thought is, any conceptual work on top of that would be superfluous.

It would be a mistake to think that the proposed task for philosophy of mind would make it into a detached discipline, devoid of interest to any other. For it would still be concerned with making the outline of concepts sharper, which should be helpful with puzzles about borderline cases. When we are concerned with a moral or judicial question about a certain situation involving unclarity

about whether an action was performed (or about the authorship of an action), it should be helpful to understand what other mental concepts should have applied for a particular scenario to form an accurate conceptual description of what happened. For example, in an inquiry into the death of Mr. Jones, who tragically died in a street-fight, what is of interest is not so much what caused the bodily movements of the people in his vicinity at the time, but rather what they saw, heard, believed, and intended. If philosophy of mind is to be of any use to any other discipline, the use of mental concepts within the mental vocabulary needs to be clarified, instead of attempting to forge links with 'more scientific disciplines'.

6.3.2 Psychological explanation

Against the idea that action explanation must be causal explanation, I have given various arguments. Especially what appears to be Davidson's view, namely that (intentional) actions distinguish themselves from other events by being liable to causal explanation in terms of reasons, has been shown to be unattractive.

I suspect that thesis **FPCE** is supposed to help put up some resistance to the eliminativist. The eliminativist, recall, argues that the mental concepts which we use in our psychological explanations will turn out to have empty reference: in other words, that there are no such things as desires and beliefs. The thesis that these concepts figure in explanations which work very well, and moreover work so well because they are of a respectable type, makes the eliminativist conclusion look implausible. But **FPCE** is not needed for this. That action explanations work is simply a datum; that people desire and believe things can be perfectly explanatory without beliefs and desires being 'causally real'.

The arguments to do with counterfactuals supported by action explanations in the previous chapter also go some way towards showing that action explanations cannot be causal. The **FPCE** proponent, in any case, has some work to do to account for the additional counterfactuals that are supported by action explanations, but not by causal explanations.

What alternative accounts are there of action explanation? If we concede that there is a question to be answered as to why action explanations are so forceful, there seem to be two possibilities. The first is the Wittgensteinian position that reasons explain actions by placing them against a certain background. As Davidson argues, that view seems to leave unexplained why the action was taken in the first place. The second possibility is to say that action explanations are teleological explanations. That position would need some filling in: what exactly

is meant by 'teleological explanation', and why does it not reduce to causal explanation? But the account does seem promising, in holding the possibility of combining Davidson's requirement that an occurrence is explained, while shifting away from an internalist representationalist view of the mental.

The third possibility would be to reject the question. Action explanations, we might hold, are *sui generis*. Why should we be held to explain what accounts for their force? Explanation in terms of reasons obviously is a very successful kind of explanation; and arguably, it is the kind of explanation we are most familiar with – plausibly more so than with causal explanation. Thinking of reason explanation as mysterious if not causal, as Davidson suggests[122], seems to me philosophers' indoctrination.

6.4 Conclusions

I have in this thesis attempted to disentangle some of the causal claims about the mental that one finds in the literature. In an extended discussion of the concepts of action and perception I have argued against two claims in particular. The first is the claim that a certain causal ancestry is a necessary condition for application of the concept in question; none of the arguments I surveyed in favour of that claim established it conclusively. The second is a claim that reason explanation must be causal explanation if we are to account for its force. When disentangled from the ancestry claim it is hard to see why this would be a distinctively causal proposal. And there are reasons to think that the real force of reason explanations is not accounted for at all in this way.

I have suggested that given the analogies between action and perception it is plausible to think that the points I have made will generalise, at least to some extent. On the other hand, to expect too much from this move would be to fall into the same kind of trap as the causalist who tries to paint a unified picture of the mind. There are, and remain, differences between mental concepts.

I will end by stating two things that I have not argued. Firstly, I have not established that causal ancestry claims must be false, only that they have not been conclusively established. Secondly, I have not established any conclusions about the subject matter of mental concepts, except maybe that the discussion about the concepts left open that they may well be states that are causally active and acted upon.

[122] Davidson 1980 p.11

7 Literature

Alston, W.P. (1993): *The Reliability of Sense Perception*, Cornell Univ. Press

Anscombe, G.E.M. (1957): *Intention*, Oxford: Blackwell

Anscombe, G.E.M. (1971): "Causality and Determination", reprinted in Sosa & Tooley (eds) 1993

Antoniol, L. (1994): "Describing Actions: with colours but without accordion", manuscript

Armstrong, D. (1968): *A Materialist Theory of the Mind*, London: Routledge

Armstrong, D. (1973): *Belief, Truth, and Knowledge*, Cambridge: CUP

Austin, J.L. (1962): *Sense and Sensibilia*, Oxford: Oxford University Press

Bach, K. (1987): *Thought and Reference*, Oxford: OUP

Baker, L.R. (1995): *Explaining Attitudes*, Cambridge: CUP

Bambrough, R. (1960): "Wittgenstein on family resemblances", repr. in Pitcher (ed.) 1968

Bishop, J. (1989): *Natural Agency*, Cambridge: CUP

Brakel, J. van & Raven D. (eds.) (1991): *Realisme en Waarheid*, Assen: van Gorcum

Bransen, J. & Cuypers S. (eds.) (1998): *Human Action and Deliberation*, Dordrecht: Kluwer

Budd, M. (1989): *Wittgenstein's Philosophy of Psychology*, London: Routledge

Cartwright N. (1983): *How the laws of physics lie*, Oxford: Clarendon

Chan, D. (1995): "Non-intentional Actions", in: *American Philosophical Quarterly*, Volume 32, no.2

Charles, D. and Lennon, K. (eds) (1992): *Reduction, Explanation and Realism*, Oxford: Clarendon

Child, W. (1994): *Causality, Interpretation, and the Mind*, Oxford: Clarendon

Child, W. (1992): "Vision and Experience: the Causal Theory and the Disjunctive Conception", in: *Philosophical Quarterly*, vol.42 no.168

Chisholm, R.W. (1957): *Perceiving: a philosophical study*, Cornell Univ. Press

Chisholm, R.W. (1966): *Theory of Knowledge*, Prentice-Hall

Churchland, P.M. (1984): *Matter and Consciousness*, Cambridge Mass.: MIT Press

Clark, A. (1989): *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing*, Cambridge (Mass):MIT

Collins, A. (1984): "Action, Causality, and Teleology", in: French & Uehling (eds.) 1984

Collins, Arthur (1987): *The Nature of Mental Things*, University of Notre Dame Press

Crane, T. (ed.) (1992): *The Contents of Experience: Essays on Perception*, Cambridge: Cambridge University Press

Crane, T. (1995a): "Mental Causation", in: *Aristotelian Society Supplementary Volume* pp.211-236

Crane, T. (1995b): *The Mechanical Mind*, Penguin

Crane, T. (1998): "A functionalist theory of content", in Bransen & Cuypers (eds.) 1998

Cross, C.B. (1992): "Counterfactuals and event causation", in: *Australasian Journal for Philosophy*, vol 70 pp.307-323

Dancy, J. (1985): *Contemporary Epistemology*, Oxford: Blackwell

Dancy, J. (ed.) (1988): *Perceptual Knowledge*, Oxford: OUP

Danto, A.C. (1973): *Analytic Philosophy of Action*, Cambridge: CUP

Davidson, D. (1980): *Essays on Actions and Events*, Oxford: Clarendon Press

Davidson, D. (1993): "Thinking Causes", in Heil and Mele (eds.) 1993

Davis, S. (ed.) (1983): *Causal Theories of Mind : Action, Knowledge, Memory, Perception and Reference*, Berlin: De Gruyter

Devitt, M.: "A Narrow Representational Theory of the Mind", in: Lycan (ed.) 1990, pp. 371-398

Dretske, F. (1988): *Explaining Behaviour*, Cambridge Mass: MIT Press

Dretske, F. (1981): *Knowledge and the Flow of Information*, Cambridge Mass: MIT Press

Dupré, J. (1993): *The Disorder of Things*, Cambridge Mass: Harvard University Press

Evans, G. (1982): *Varieties of Reference*, Oxford: OUP

Fodor, J. (1974): "Special Sciences", reprinted in Fodor 1981

Fodor, J. (1981): *RePresentations*, Brighton: Harvester Press

Fodor, J. (1987): *Psychosemantics*, Cambridge Mass: Bradford Books

Fraassen van, B.C. (1980): *The Scientific Image*, Oxford: Clarendon

French, P.A., Uehling, T.E. (jr.), Wettstein, H.K. (eds.) (1984): *Midwest Studies in Philosophy IX: Causation and Causal Theories.*

Garcia-Carpintero, M. (1995): "The philosophical import of connectionism", in: Mind and Language Vol 10 no.4 pp. 370-401

Gardner, H. (1985): *The Mind's New Science*, New York: Basic Books

Gettier, E. (1963): "Is justified true belief knowledge?", in *Analysis* 23, reprinted in Griffiths (ed.) 1967

Goldman, A.I. (1967): "A Causal Theory of Knowing", in: *Journal of Philosophy* 64

Goldman, A.I. (1976): "Discrimination and perceptual knowledge", in: *Journal of Philosophy* 73

Goldman, A.I. (1977): "Perceptual Objects", in: *Synthese* vol.35 pp.257-284, reprinted in Davis (ed) 1983

Grice, P. (1961): "The Causal Theory of Perception", *The Aristotelian Society Proceedings*, Supp. vol. 35, reprinted in Dancy (ed) 1988

Grice, P. and Strawson, P. (1956): "In defence of a dogma", in *Phil Review* pp.141-158

Griffiths, A.P.(ed.) (1967): *Knowledge and Belief*, Oxford: OUP

Guttenplan, S.(ed.) (1994): *A Companion to the Philosophy of Mind*, Oxford: Blackwell

Hale, B. (1987): *Abstract Objects*, Oxford: Blackwell

Harré, R. and Madden, E.H. (1975): *Causal Powers*, Oxford: Blackwell

Heal, J. (1994): "Moore's Paradox: a Wittgensteinian Approach", in *Mind* vol 103 409 Heil, J. and Mele, A. (eds) (1993): *Mental Causation*, Oxford: OUP

Hinton, J.M. (1967): "Visual Experiences", in *Mind* 76

Hornsby, J. (1980): *Actions*, London: Routledge

Hume, D. (1902): *An Enquiry Concerning Human Understanding*, L.A. Selby-Bigge (ed), Oxford: Clarendon

Hursthouse, R. (1991): "Arational Actions", in: *Journal of Philosophy* vol 88

Hyman, J. (1992): "The Causal Theory Of Perception", in: *Philosophical Quarterly*, vol.42 no.168

Jackson, F. (1977): *Perception: a representative theory*, Cambridge: CUP

Jackson, F. (1982): "Epiphenomenal Qualia", in: *Philosophical Quarterly* 32 pp. 127-36

Kazez, J. (1995): "Can Counterfactuals Save Mental Causation?", in *Australasian Journal of Philosophy*

Kim, J. (1988): "Explanatory Realism, Causal Realism and Explanatory Exclusion", in: French, Uehling & Wettstein (eds) *Midwest Studies in Philosophy XII*

Kim, J. (1973): "Counterfactuals and causation", in: *Journal of Philosophy* vol.70

Kim, J. (1993a): "Can Supervenience and 'Non-Strict Laws' Save Anomalous Monism?", in Heil and Mele (eds.) 1993

Kim, J. (1993b): "Mental Causation and two Conceptions of Mental Properties", paper presented at XVIth Ludwig Wittgenstein Symposium in Kirchberg, Austria

Kim, J. (1993c): "The Non-Reductivist's Troubles with Mental Causation", in Heil and Mele (eds.) 1993

Klein, M. (1996): "Externalism, content and causation", in: *Proceedings of the Aristotelian Society* XCVI pp.159-171

Kripke, S. (1980): *Naming and Necessity*, Harvard University Press

Lehrer, K. (1989): *Thomas Reid*, London: Routledge

Lennon, K. (1990): *Explaining Human Action*, London: Duckworth

LePore, E. and McLaughlin, B.P. (1985), *Essays on Actions and Events*, Oxford: Blackwell

Lewis, D. (1973): "Causation", in: *Journal of Philosophy* vol. 70 pp.556-567

Locke, J. (1975): *An Essay Concerning Human Understanding*, ed. P. Nidditch, Oxford: Clarendon

Lowe, E.J. (1995): *Locke on Human Understanding*, London: Routledge

Lycan, W.G.(ed.) (1990): *Mind and Cognition*, Oxford: Blackwell

Lyons, Wm.: "Intentionality and modern philosophy of psychology, part III: The appeal to teleology", in *Philosophical Psychology* (1993)

Macdonald, C.I. (1989): *Mind-Body Identity Theories*, Oxford: Blackwell

Macdonald, C.I. and Macdonald, G (eds.) (1995): *Philosophy of Psychology*, Oxford: Blackwell

Malcolm, N. (1968): "The Conceivability of Mechanism", in: *Physical Review* 77

Malcolm, N. (1963): *Knowledge and Certainty*, Cornell Univ. Press

Malcolm, N. (1977): *Memory and Mind*, Cornell Univ. Press

Manser, A. (1967): "Games and Family Resemblances", in *Mind* 76

Marr, D. (1982): *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, San Francisco: Freeman

McCulloch, G. (1995): *The Mind and its World*, London: Routledge

McDowell, J. (1982): "Criteria, Defeasibility and Knowledge", repr. in Dancy (ed) 1988, pp.209-219

McDowell, J. (1986): "Singular Thought and the Extent of Inner Space", in: Pettit and McDowell (eds.) 1986

McDowell, J. (1994): *Mind and World*, Harvard Univ. Press

McLaughlin, B.P. (1984): :"Perception, Causation, and Supervenience", in French & Uehling (eds): *Midwest Studies in Philosophy IX: "Causation and Causal Theories"*

McLaughlin, B.P. (1993):"On Davidson's response to the Charge of Epiphenomenalism", in Heil and Mele (eds) 1993

Melden, A.I. (1961): *Free Action*, London: Routledge

Mellor, D.H. (1991): *Matters of Metaphysics*, Cambridge: CUP

Mellor, D.H. (1995): *The Facts of Causation*, London: Routledge

Melzack, R. (1997): "Phantom Limbs", in *Scientific American* vol.7 no.1

Merleau-Ponty, M. (1981): *Phenomenology of Perception*, transl. Colin Smith, (rev. ed.) London: Routledge

Michotte (1963): *The Perception of Causality*, London: Methuen

Millar, A. (1991): *Reasons and Experience*, Oxford: Clarendon

Millar, A. (1995): "The Idea of Experience", in: *Proceedings of the Aristotelian Society* pp. 75-90

Millikan, R. (1984): *Language, Thought, and other Biological Categories*, Cambridge Mass.: MIT Press

Moya, C.J. (1990): *The Philosophy of Action*, Cambridge: Polity Press

Nagel, T. (1979): "What is it like to be a bat?", in: *Mortal Questions*, Cambridge: CUP

Nagel, T. (1986): *The View from Nowhere*, Oxford: OUP

Neuberg, M. (1993): *Philosophie de l'action*, Bruxelles: Academie Royale de Belgique

Nozick, R. (1981): *Philosophical Explanations*, Oxford: OUP

Oppenheim & Putnam (1958): "Unity of Science as a Working Hypothesis", in: Feigl, Scriven & Maxwell (eds.), *Concepts, Theories and the Mind-Body Problem*, Minnesota Studies in the Philosophy of Science vol 2

Owens, D. (1992): *Causes and Coincidences*, Cambridge: CUP

Papineau, D.: "Irreducibility and Teleology", in Charles and Lennon (eds.)

Papineau, D. (1993): *Philosophical Naturalism*, Oxford: Blackwell

Pears, D. (1976): "The Causal Theory of Perception", in *Synthese* 33, pp 41-74

Pettit, P. and McDowell, J. (eds.) (1986): *Subject, Thought and Context*, Oxford: Clarendon Press

Pitcher, G. (ed) (1968): *Wittgenstein - the Philosophical Investigations*, New York: MacMillan

Putnam, H. (1967): "The nature of mental states", repr. in: Lycan (ed) 1990 pp.47-56

Putnam, H. (1975): "The Meaning of 'Meaning' ", in K. Gunderson (ed) *Minnesota Studies in the Philosophy of Science* vol.7, Minneapolis: University of Minnesota Press

Putnam, H. (1988): *Representation and Reality*, Cambridge Mass: MIT Press

Quine, W.V.O. (1953): "Two Dogmas of Empiricism", in: *From a logical point of view* , pp.20-46, Cambridge Mass: Harper

Quine, W.V.O. (1969): "Epistemology Naturalized", in: *Ontological Relativity and Other Essays*, New York: Columbia University Press

Quine, W.V.O.: "Natural Kinds", in: *Essays in honour of Carl G. Hempel*, ed. N. Rescher, pp.5-23, Dordrecht: Reidel

Reid, Th.: *Essay on the Active Powers of Man*

Robinson, H. (1990):"The Objects of Perceptual Experience", in *Proceedings of the Aristotelian Society*, Supp.Vol.LXIV

Robinson, H. (1994): *Perception*, London: Routledge

Russell, B. (1967): *The Problems of Philosophy*, Oxford: OUP

Ryle, G. (1949): *The Concept of Mind*, London: Hutchinson

Ryle, G. (1954): "Perception", in: *Dilemmas*, Cambridge: CUP

Searle, J. (1984): *Minds, Brains, and Science*, Cambridge Mass.: Harvard Univ. Press

Smith, P. (1992):"On:'The Objects of Perceptual Experience'", *Proceedings of the Aristotelian Society*

Smith, P. & Jones, O.R. (1986): *The Philosophy of Mind*, Cambridge: CUP

Snowdon, P. (1981): "Perception, Vision, and Causation", repr. in: Dancy (ed) 1988

Snowdon, P. (1990): "The Objects of Perceptual Experience", *in Proceedings of the Aristotelian Society*, Supp.Vol.LXIV

Sober, E.: "Putting the function back into functionalism", in: Lycan (ed) 1990

Sosa, E. & Tooley, M. (eds.) (1993): *Causation*, Oxford: OUP

Stalley, R.F. (1989): "Causality and Agency in the Philosophy of Thomas Reid", in: M. Dalgarno and E. Matthews (eds.): *The Philosophy of Thomas Reid*, Dordrecht: Kluwer

Sterelny, K. (1990): *The Representational Theory of Mind*, Oxford: Blackwell Steward, H. (1997): *The Ontology of Mind*, Oxford: Clarendon

Stout, R. (1996): *Things that happen because they should*, Oxford: Clarendon

Stoutland, F. (1976): "The Causation of Behaviour", in: "Essays on Wittgenstein in honour of GH von Wright", Acta Philosophica Fennica XXVIII

Stoutland, F. (1980a): "Davidson, von Wright, and the debate over causation", in *Contemporary Philosophy*, ed. Floistad

Stoutland, F. (1980b): "Oblique Causation and Reasons for Action", *Synthese* vol.43.

Stoutland, F. (1982): "Davidson on Intentional Behaviour", in Lepore & McLaughlin (eds)

Stoutland, F. (1998): "The Real Reasons", in: Bransen & Cuypers (eds.)

Strawson, P.F. (1974): "Causation in Perception", in his *Freedom and Resentment and other essays*

Strawson, P.F. (1979): "Perception and its Objects", repr. in: Dancy (ed) 1988

Strawson, P.F. (1985): "Causation and Explanation", in Vermazen & Hintikka (eds.)

Taylor, Charles (1964): *The Explanation of Behaviour*, London: Routledge

Taylor, Richard (1966): *Action and Purpose*, Prentice-Hall

Travis, C. (1989): *The Uses of Sense*, Oxford: OUP

Tuomela and Manninnen (eds.): *Essays on Explanation and Understanding*

Tye, M. (1993): *The Metaphysics of Mind*, Cambridge: CUP

Vermazen and Hintikka (eds) (1985): *Essays on Davidson*, Oxford: Clarendon

Wilson, George M. (1989): *The Intentionality in Human Action*, Stanford University Press

Wilkie, S. (1996): "The Causal Theory of Veridical Hallucinations", in: *Philosophy*

Williamson, T. (1995): "Is Knowing a State of Mind?", in: *Mind* vol 104, 415

Wittgenstein, L.: (1953) *Philosophical Investigations*, Oxford: Blackwell

Wittgenstein, L.: *Remarks on the Philosophy of Psychology*, vol I, Oxford: Blackwell

Woodfield, A. (1976): *Teleology*, Cambridge: CUP

Woodfield, A. (ed) (1982): *Thought and Object. Essays on Intentionality*, Oxford: Clarendon Press

Wright, L. (1976): *Teleological Explanations*, Univ. of California Press

Wright von, G.H. (1971): *Explanation and Understanding*, New York: Cornell University Press

Wright, C.G. (1986): "Inventing logical necessity", in: Butterfield (ed.), *Language, Mind, and Logic*, Cambridge: CUP

Wright, C.G. (1996): "Judgement dependence and anomaly", manuscript

Yablo, S. (1992): "Mental Causation", in *Philosophical Review* 101 no.2