MANIPULATING COLOUR

BY JOHN CAMPBELL

I want to propose that in many cases, understanding a concept is a matter of knowing the causal role of a property. And this notion of 'knowing the causal role of a property' can be explained in manipulationist terms. That is, it can be explained in terms of the idea that for C to cause E is for manipulating C to be a way of manipulating E.

In this essay, I will illustrate this general approach by looking in particular at the case of colour concepts. Colour concepts raise very sharply the problem of the role of experience in understanding, for it seems compelling that to understand colour concepts you must have colour experience. I will propose that the role of colour experience is to make it possible for us to attend to colours for the purpose of manipulating them; I will argue that this does not involve our making any mistake about the causal processes at work here.

1. The Commonsense Concept of Colour

Locke said that science shows that there is a mistake embodied in our ordinary understanding of colour concepts. We commonsensically take colours to be categorical properties of objects, whose nature is apparent to us in vision, but in fact there are only complex microphysical structures and the consequent tendencies of objects to produce

ideas in us. Those who have followed Locke in holding that there are only the microphysical structures and the tendencies to produce experiences in us have often also agreed that there is an error that we naively fall into here: that of supposing that colours are categorical or intrinsic properties of objects, displayed to us in vision. Even if, like Locke, you think that the naïve conception is mistaken, it does seem to be the conception of colour that we have pre-scientifically. And we ought to be able to explain how it is that we have this conception of colour as categorical.

Even for properties, such as colour, that we can easily observe objects to have, we do not usually suppose that an object having the property is the same thing as us having certain perceptions. We usually take it that perception of an object as having the property is due to the operation of two different sorts of factor: the object having the property, and the perceiver being appropriately positioned, looking in the right way, and so on, with respect to the object. Whether the object is red is one thing, and whether I am so positioned as to be able to see that it is, is another thing. In consequence of this, we ordinarily take it that there is a difference between changing the colour an object intrinsically has, and changing the way it looks to an observer, by manipulating the conditions of perception. And I want to propose that grasp of this distinction is central to our ordinary understanding of colour concepts. More generally, what I am proposing in general is that to know what it is for an object to have the property is to know something about the causal role of the property itself.

What is the role of experience in providing us with this conception of colour as a categorical, or intrinsic property of objects? On a classical semantic theory, a name makes its contribution to the meaning of a sentence by standing for an object. And a

predicate makes its contribution to the truth or falsity of a sentence containing it by standing for, in Michael Dummett's phrase, a mapping from objects to truth-values. Understanding a predicate such as 'is red' is a matter of knowing which mapping from objects to truth-values is associated with the predicate. That is what it is, to know which property the predicate stands for. So on a classical semantic approach, we will take it that the role of colour experience is to provide knowledge of what it is for an object to have the colour property.

There is a quite different account you might give of the role of experience in understanding colour concepts. On this account, the role of experience in understanding colour concepts is to provide us with reasons for making judgements of colour. Learning the colour concept, on this view, is a matter of learning which experiences constitute reasons for making judgements in which the concept is applied to an object. There is nothing more fundamental, in grasp of a colour concept, than knowledge of which experiences constitute reasons for making which colour judgements.

There is something instrumental about the notion of a 'reason'. That is, a reason is always a reason-for something. So, on the face of it, for colour experiences to provide reasons for colour judgements, there must be such a thing as grasp of the truth-conditions of those judgements. That grasp of truth-condition will provide an understanding of what one is aiming at in verifying a colour proposition. The problem then for the reasons-based approach is to explain how we can have this conception of the truth-condition of a colour judgement, and what the role, if any, might be of experience in providing one with such a conception. On a reasons-based approach, the role of the colour experience can't be directly to provide knowledge of what it is for an object to have a particular colour.

That is just what is meant by saying that what is fundamental is the role of experiences in providing reasons, rather knowledge of truth-conditions.

The reasons-based theorist might argue that colour experience has a further role to play, over and above providing a reason for making a judgement about the colour of a seen object. Colour experience also plays a role in providing the subject with the conceptions of particular types of colour experience. And once we know what it is for someone to have a particular type of colour experience, we can form the conception of an object's having a tendency to produce that type of experience. And if you have an experience of that type, that may of itself prompt the hypothesis that the object you perceive has a tendency to produce that type of experience. So a 'reasons-based' approach may propose that the natural conjecture for us to form about the truth-conditions of colour judgements is the dispositionalist one. On the dispositionalist account, the truth of a colour judgement depends on whether the object has a tendency to produce the right type of colour experiences in us.

One problem here is that the conjecture which the subject is supposed to make, deriving the truth-condition from knowledge of what constitutes a reason for making a colour judgement, does not seem defensible. From the fact that colour experiences provide reasons for making colour judgements, all the subject can conclude is that having a particular experience is correlated with an object's having a particular colour. So all the subject can say is that for an object to have a particular colour is for it to have some property that is correlated with the occurrence of these experiences. This could happen in a number of ways, even if we think there could be no such thing as 'spontaneous correlation' with no causal explanation. There might be some third property that causes

the object to have the colour and also causes the perceiver to have the experiences. Or the perceiver having the experience might cause the object to have the property. There is no evident reason to fasten on to the idea that the colour must be a disposition to produce that experience in the subject, given only that the experience is a reason for judging the object to have the colour.

Moreover, this reasons-based account makes error theories of colour impossible. The account is in effect arguing that we could not have the conception of colour as an intrinsic characteristic which, the error theorist says, science has shown to be mistaken. For, the reasons-based theorist is arguing, the only conception of colour we could have formed is the conception of colour as a disposition of objects to produce colour experiences in us.

It therefore seems worth pursuing the classical account further. That is, we should try to articulate the notion that the role of colour experience in understanding colour concepts is not in the first instance to provide one with knowledge of what constitutes a reason for making a colour judgement. Nor is the role of experience to provide one with the concept of colour experience itself. Rather, the role of colour experience is to provide one with knowledge of which categorical properties the colours are. Such an account will explain how it is that we have the conception of colour that error theories attack.

Earlier I suggested that on this common-sense conception, we think of perception of the colour of an object as the joint upshot of two factors: the intrinsic, or categorical colour property that the object has, and the facts about the positioning of the observer with respect to the object ; it being in a good light, the observer looking in the right

direction, and so on. So we think of the colour of the object as one of the causes of perception having the content it does. On a manipulationist account of causation, that is to say that an intervention on the colour of the object would change the content of the experience. So we could think of knowledge of the colour properties as knowledge of what the upshot would be of various potential interventions on the colours of objects. On this account, the colour of the object is not simply a disposition the thing has to produce particular experiences in you. Rather, the colour of the object is the property in virtue of which the thing has that disposition. And experience confronts you with that property. But experience confronts you with that property by confronting you with its causal significance. To fill out this approach, let me begin by saying something about the manipulationist notion of causation to which I am appealing.

2. Causation

Manipulationism is the idea that for C to cause E is for manipulating C to be a way of manipulating E; it has been developed in the work of Sprites, Glymour and Scheines (1993), Pearl (2000) and given its fullest statement in Woodward (2003).

A simple way to represent causal relations among a family of variables is a causal graph, which draws an arrow between two variables when one is a direct cause of another. In contrast to a set of equations, a causal graph does not distinguish between promoting and inhibiting causal factors. Moreover, when two or more variables affect an outcome variable, the graph does not distinguish between the case in which the input

variables interact with one another to affect the outcome, and the case in which the input factors have their impact on the outcome variable separately. In this essay I'll be concerned with causal relations between variables, such as income or longevity, which can take many different values; people have different incomes and live for varying amounts of time. To say that income is a cause of longevity is to say that the particular income any individual has causally affects just how long that individual lives.

Manipulationism says that we can explain what it is for X to cause Y by saying that an intervention on the value of X would make a difference to the value of Y, in some cases at any rate; it need not be that every change in X would make a difference to Y. The idea is that we can define an 'intervention' on the value of X so that any change in the value of Y can only be due to a causal connection between X and Y. But to avoid unilluminating circularity, we have to try to give the definition without using the notion of a causal connection between X and Y.

Judea Pearl explained an intervention as being a surgical disruption of an existing causal mechanism. The idea is that you lift X from the influence of the mechanism that usually sets its value, and place it under the influence of a new mechanism which now sets the value of X, while leaving the remainder of the causal mechanism undisturbed. In that case, the proposal is, changes in the value of any variable other than X will have to happen in virtue of a causal relation between X and that variable.

Woodward and Hitchcock give a more explicit definition in similar vein. They say that an intervention variable I for X with respect to Y is one which:

1) I is causally relevant to X.

> 2) I is not causally relevant to Y through a route that excludes X.
>
> 3) I is not correlated with any variable Z that is causally relevant to Y through a route that excludes X, be the correlation due to I's being causally relevant to Z, Z's being causally relevant to I, I and Z sharing a common cause, or some other reason.
>
> 4) I acts as a switch for other variables that are causally relevant to X. That is, certain values of I are such that when I attains those values, X ceases to depend upon the values of other variables that are causally relevant to X.
>
> (Woodward and Hitchcock 2003,)

The reason I quote this explicit definition is partly to make it evident that the notion of an intervention can be defined without any reference being made to human action. Any external influence that meets the conditions I described will do. Nonetheless, it does not seem to be an accident that it is so intuitive to use the notion of 'manipulation' here. The kinds of experimental intervention characteristic of science, to determine causal influences, seem to take it that actions by the experimenter will aim at being interventions in the sense characterised above. And in a number of recent studies, Gopnik and her collaborators (see, e.g., Gopnik et. al. in press) have argued that very young children, in exploring the causal structure of their surroundings, take it that their own free actions are paradigmatic interventions in the sense described. And indeed, given that this account of causation does not attempt to provide a reductive analysis, there is a question as to where we begin in establishing the existence of causal relations. The natural proposal is in

many cases, we begin with the idea our own manipulations of the world constitute interventions in the sense just explained.

My question was, what is it to know the causal role of colour properties? In particular, what is the role of experience in providing us with knowledge of the causal roles of colour properties? In our present terms, this becomes the question, what is it to know what the outcome would be of interventions on colour properties?

3. Conscious Attention to Colours

Let us go over the idea that experience of the colours plays a role in our understanding of colour concepts. Someone who is blind or entirely colour-blind from birth, or someone who is normally sighted but simply never encounters colours, cannot understand colour predicates as we ordinarily understand them. Experience of the colours does some work in our ordinary grasp of colour concepts. Still, the kind of colour experience needed needs careful explanation. Recall the kind of test for colour vision that consists of a number of variously coloured dots of various sizes in a single display. The colouring of the dots may be so organised that, let us suppose, someone with ordinary colour vision can quite plainly see a figure, say the numeral 5, picked out in some one colour, say gold. For anyone without colour vision, though, all that they can see is an array of variously shaded and variously sized dots. So an ability to identify that there is a number 5 in the array provides good evidence that the subject has ordinary colour vision. Someone who can see the figure 5 in this kind of display need not, however, be capable of visually

attending to the colours of things; they may not realise that there is such a thing as colour at all. This subject is attending only to the object, the number 5, not to the characteristics which allowed him to discriminate the object. Such a person might, for all that I have said so far, be unable to report the colours of objects, or to match different objects which are the same colour.

For experience to provide knowledge of the colours, the subject must not only have colour experience, but must be capable of visually attending to the colours of the things he sees. There are, though, many tasks that involve attention to colours which do not seem involve an understanding of colour concepts. Suppose we have a subject who performs the following tasks. When given two rows of coloured paper, he can match each paper in one row to the same-coloured paper in the other row. Or again, when he is given a pile of chips of two slightly different shades of green, he sorts them successfully. He can correctly arrange a series of reds in order from bright red at one end to pink at the other. Given coloured papers or crayons and a group of line drawings of familiar objects, he could correctly match the colours to the objects, for example, the yellow crayon to the banana. In performing these tasks he is plainly attending to the perceived colours of objects. Let us suppose that he also passes the tests I mentioned earlier: he can use colour as an object-defining property. He performs well on the American Optical Company and the Ishihara pseudo-isochromatic tests of colour vision – that is, discerning the figure 5 in a pattern of blobs, and so on.

Geschwind and Fusillo 1997 describe a patient, 58 years old, whose performance in tasks of colour identification was as I have just described. However, when he was asked to name the colour of a figure shown, his replies were wildly inaccurate. For

example, a card which showed a bright red 7 on a grey background was described as having a grey 7. When the patient was shown an array of variously coloured objects, such as several sheets of paper, and asked to 'show me the red one' for example, he usually failed; he also answered at chance when shown a sheet of paper and asked, 'Is this red?' When presented with coloured sheets of paper, and asked to name their colours, he gave incorrect answers in almost all cases, including cases in which he was presented with sheets of black, white or grey paper. When the patient was presented with colour pictures of objects such as neckties or curtains, which can be of any of a variety of colours, and asked to name their colours, he made similar errors. When shown coloured pictures of objects such as bananas or milk, which have standard colours, and asked to name those colours, the patient again was almost invariably wrong.

It is not, however, as though there were specifically verbal problem here. The patient could identify the objects verbally, as bananas or milk and so on,. And when asked the usual colours of objects such as bananas or milk, the patient performed without error. When asked to give examples of objects which standardly have a certain colour, he again performed without error. Geschwind and Fusillo comment:

> The patient failed in all tasks in which he was required to match the seen colour with its spoken name. Thus, the patient failed to give the names of colours and failed to choose a colour in response to its name. By contrast, he succeeded in all tasks where the matching was either purely verbal or purely nonverbal. Thus, he could give verbally the names of colours corresponding to named objects and vice versa. He could match seen colours to each other and to pictures of objects

> and could colours without error. By no nonverbal criterion could our patient be shown to have any deficit in colour vision.
>
> (Geschwind and Fusillo 1997, p.271)

Suppose we extrapolate somewhat from the Geschwind and Fusillo results. Suppose that this patient is successful in all purely verbal tests of knowledge of the colours. That is, he knows, for example, that nothing can be both green and red all over. He can verbally order the colours, can say that orange is between yellow and red, and so on. And he also passes all the purely non-verbal tests for colour vision. His problems come only with the liaisons between colour names and colour vision. Can this patient be said to grasp colour concepts?

A reasons-based theorist has a simple diagnosis of the situation here: this patient has lost his grasp of colour concepts, because he has lost his ability to recognise which colour experiences are reasons for which judgements of colour. But a different diagnosis is possible. This would run as follows.

There are many different modulations of attention to colour. You may be attending to colour for any of endlessly many different purposes. And we do not yet know enough about this patient's attention-based skills to determine whether he has the basis for grasp of colour concepts. In particular, we have to look at whether the patient can put attention to the colours to work in a way that displays grasp of the causal roles of the colours.

You might object that the only capacity we need to look at is the subject's capacity to say, or to judge, when the object has some particular colour. This view seems

to be implicit in many characterisations of the role of colour experience in our grasp of colour concepts: the only role envisaged for colour experience is as making it possible for the subject to say, or to judge, when a particular seen object has a particular colour. On this approach, all I have done so far is to point out that the capacity for attention specifically to the colour of the object is demanded by the ability to make such statements or judgments of colour. What I shall try to show now, though, is that there is more to our ordinary grasp of colour concepts than the mere ability to apply colour predicates to seen objects.

We have so far no explanation of the sense in which the subject can be said to know the causal role of colour properties. And there is a richer account available to us. There is more to learning colour concepts than merely learning how to give reports of the colours of objects in response to observation of them. We also have to learn something about the causal role of colour. On the account I sketched earlier, learning about the causal role of colour is learning something about what the outcome would be if there were to be an intervention affecting the colour of an object. But here colour has some quite special characteristics. In the case of shape, for example, there are many purposes to which we can put manipulation of shape. You might bend something to fit it through a letterbox. You may want to manipulate the shapes of things to roll them, to stack them together for easy carrying, to wrap them around you, to use them as tools. But colour does not have the same broad causal significance as shape. Of course, colour is often symptomatic of the further characteristics of an object. This is particularly so for children living in present-day environments full of colour-coded toys. But even in the wild, colour is important for pursuits like finding good food, or deep water. But you

can't, in general, change the further characteristics of an object by changing its colour. It would be very unusual for it even to occur to someone to try to affect whether or not the food was good to eat by manipulating its colour. There is no analogue, for colour, of trying to get the envelope through the letterbox by manipulating its shape. The exception is, of course, that by manipulating the colours of objects you can make a difference to the experiences that people will have when they look at those objects. This is what you have to grasp if you are to grasp the causal roles of the colours.

You could in principle learn purely from observation about what would happen to the experiences of observers if the colour of an object were manipulated. You could extract this information just by finding the probabilistic relations between the colours of the objects and the experiences of the observers. But it is not evident what the point would be of doing that, unless you were yourself capable of manipulating the colour of an object. Children's slowness in learning a colour vocabulary does seem baffling so long as we think that there is no more to learning a concept than learning to use a word in response to observation. But I am proposing that we should see the use of words in response to observation as a phenomenon that, in many cases, is an offshoot of grasp of the causal significance of a property. The enthusiasm children have for finding a large box of crayons and manipulating the colours of everything in sight is not an accompaniment to their grasp of colour concepts; it is fundamental to it. So the idea I am presenting here could be put as the idea that learning to manipulate the colours of objects, using for example crayons, is the way in which you typically come to grasp of the causal role of colour properties; and this in turn is what constitutes your grasp of the colour concept.

On this richer account, we have to look at the role of conscious attention to colour in connection with the subject's grasp of what would happen to the values of other variables, were the colour of an object to be manipulated.

This approach brings out something of why it matters, for grasp of colour concepts, that we have experience of the colours, rather than merely brain states which are differentially sensitive to the various colours. Experience of the colours means that you know which variable you are manipulating, in a way that you would not if you had only brain states which were differentially sensitive to the colours. Just to make this point vivid, suppose we consider someone whose conscious experience is all in black and white. But his environment contains coloured objects – the things in front of him, the things he sees, are actually coloured. It's just that he cannot see their colours. There has been some damage to his visual system so that his experience is entirely in black and white. Nonetheless, suppose we ask him to guess at the colours of the objects around him. At first, of course, he says that he doesn't know what colours any of them have. But we ask him to guess. And he reliably gets it right about the colours of the objects around him – let's assume that he knows the names of all the colours, so he knows what would be an appropriate answer to the question, 'What colour is that?' And of course he does have experience of the individual objects being pointed out, so he has no problem about knowing which object is in question. This subject could in principle be just as reliable as anyone else in verifying judgements about the colours of the objects around him. There is certainly a sense in which this subject can be said to be attending to the colours of the objects he sees. He has, after all, at any one time a welter of information from the various senses. Some of this welter of information is being selected and used to

control which verbal reports he is making about the things around him. And his verbal reports are accurate reports specifically of the colours of things. In fact, they might, consistently with my description of him so far, depend on the very same neural cell-firings as do our ordinary judgements of colour. So this subject has states which, on the face of it, do the work that a reasons-based theorist asks of colour experiences in verifying judgements of colour. But when you ask this subject to act on the colours of the things being perceived by himself and others, by changing the colours of the objects before him, it is evident that this subject would not know which variable he was manipulating. Experience of the colour is an essential element in your knowledge of which variable you are manipulating. And it is this knowledge that constitutes grasp of the colour concepts.

4. Pounding an Almond

I think that the simplest way to interpret the error theorist is as accepting something like the account I have given of our ordinary colour concepts. Nonetheless, the error theorist says, it is a mistake to suppose that experience directly confronts you with the variable you are manipulating when you intervene to change the colour of an object, and thereby make a difference to the values of other variables. Here is Locke:

> Pound an Almond, and the clear white *Colour* will be altered into a dirty one, and the sweet *Taste* into an oily one. What real Alteration can the beating of the Pestle make in any Body, but an Alteration in the *Texture* of it?
>
> (*Essay*, II/viii/20)

The general question is how to characterise the variables on which you are intervening in a manipulation. The challenge is: what we take to be interventions on the colour of an object are more properly thought of interventions on the microphysical properties of the object. The point is to look at what it is that a pestle does, in general, to the object it pounds. The pestle is not in general a device that changes the colours of things. It would be kind of magic, if in the case of almonds specifically, the pestle had the capacity to change the colour of the thing directly, rather than by manipulating any other variable. Rather, the pestle does what it always does, and operates mechanically to affect shape, size and motion. It is when we regard it as affecting the shape, size and motion of atoms, Locke is saying, and only consequently affecting the colour of the almond, that we make sense of the situation.

Recall the discussion of intervention and manipulation in §2 above. I said that to be displaying the causal significance of variable X as affecting variable Y, an intervention I on X must not affect Y in any way other than by affecting X. So, for example, if we are to display the causal connection between the level of a drug in someone's body, and the speed of their recovery from an illness, the intervention that we use to vary the level of drug in the body must not affect the speed of recovery directly. That is what we are controlling for when we control for placebo effects. The giving of

the pill is how we are intervening to affect the level of drug in the body. If the giving of the pill affects the speed of recovery directly – otherwise than by affecting the level of drug in the body – then the intervention has not displayed any causal relation between the level of drug and the speed of recovery. Locke's remark is intended to raise a parallel suspicion about manipulations of the colours of objects. When we pound the almond, we change the colour of the object. There is then a change in the colour experiences of observers. But we have changed the colour of the almond only by affecting the microphysical properties of the almond. The question then is how we are to determine whether the changed microphysical properties of the almond have not affected the colour experiences of the observers directly; that is, otherwise than by affecting the presumed categorical colour of the almond.

We could put the same point another way by saying that the threat is that the microphysical facts about the almond will screen off the colour experiences of observers from the categorical colour of the almond. Learning the facts about the colour of the almond will not provide any additional information about the experiences observers will have, once we know the microphysical facts about the almond. Or again, the probability that observers will have particular colour experiences on looking at the almond, given that it has a particular microphysical constitution, is no different to the probability that observers will have a particular colour experience, on looking at the almond, given that it has a particular microphysical constitution and that it has a particular colour. The use of this kind reasoning to determine that one factor rather than another is causally relevant to an outcome is ubiquitous. The error theorist is in effect using this kind of reasoning to establish that when we take ourselves to have changed colour experiences by changing

the colour of an object, what has actually happened is that we have changed colour experiences by changing the microphysical properties of the object; the presumed change in the categorical colour of the object is an epiphenomenon.

This problem arises because when you manipulate a colour you cannot but be manipulating a physical state. And we do not yet have a way of saying what the difference is between the case in which you are manipulating a colour by manipulating an underlying physical state, and the case in which you are directly manipulating the colour and only in so doing affecting the underlying physical state.

I think the issue is whether we can find a physical variable which varies systematically with variation in colour. The mere fact that the facts about colour are entirely constituted by the physical does not of itself mean that there will in general be any systematic variation of colour variables as a result of change in physical variables. Continuous variation in an underlying physical variable might be accompanied by apparently random changes in which colour, if any, ensued.

So one response to this argument is simply to acknowledge the correctness of Locke's point for the case of changing colour by pounding, or for a wide range of similar cases, such as the use of fire to scorch and thereby change the colour of an object. In these cases effects on colour do seem, even to common sense, to be by-products of broader systematic changes brought about by this kind of intervention. In contrast, though, there is the whole broad class of paints and dyes, inks and other colourants, whose general systematic effect does seem to be to make changes in the colours of objects, even though their operation is by no means universal: black dye will not make absolutely everything black, just as pounding will not affect the shape and movement of

every object pounded. It is not an appeal to magic to propose that the use of black paint affected the object's colour directly. Of course, when the object was painted black, there will have been changes in the underlying microphysical structure of the object, on which the blackness supervenes. But that does not show that the only causality here was at the level of the supervenience base.

Whether or not common-sense makes a mistake about the world, my main aim has been to describe our ordinary understanding of colour concepts, and to explain how it can be that we have the conception of colours as intrinsic, categorical properties of objects. I said that knowledge of what the colours are is provided by the capacity for conscious attention to the colours of objects, in particular, by conscious to attention to colour for the purpose of manipulating colour to affect the experiences of observers of the object. You might wonder whether the emphasis on knowledge of the causal role of the property must not mean that there is no more to the property than its causal role; should we not then immediately equate properties with dispositions? But that goes too far. Someone who says that concrete objects are individuated by their spatiotemporal locations, and that to know which object one is talking about one must know something about its spatiotemporal location, is not thereby committed to supposing that concrete objects simply are spatiotemporal locations. Similarly, it is possible to hold that categorical properties are individuated by their causal powers, without holding that categorical properties simply are causal powers. In the case of concrete objects, you might hold that experience provides us with further knowledge of which substantial thing it is that we conceive to have a particular spatiotemporal location. Similarly, conscious attention to colour provides one with knowledge of which property one is manipulating in

varying the colour experiences of observers by manipulating the colours in a scene. And our ordinary colour concepts seem no more vulnerable to charges of error here than do our ordinary psychological concepts.

5. Systematic Manipulation and Causal Levels

I want finally to put these remarks about error theories into a broader context. There is a general problem which arises whenever we have high-level classifications which supervene on phenomena at some lower level of description. Suppose we have two high-level variables, H1 and H2, and we suppose provisionally that H1 causes H2. Then whenever we have an instance of H1 we will have an instance of some lower-lever state L1, and whenever we have an instance of H2 we will have an instance of some lower-level state L2. The general problem is to explain the distinction between the case in which the causation is a high-level phenomenon, properly described at the level of the variables H1 and H2, and the case in which the causation is a low-level phenomenon, properly described at the level of L1 and L2.

In manipulationist terms, the problem arises because when you manipulate a high-level state you cannot but be manipulating a lower-level state. When, in a particular case, we manipulate H1, we cannot but be manipulating the relevant L1. So we do not yet have a way of saying what the difference is between the case in which it is the manipulation of H1 that is causing H2, and the case in which it is the manipulation of L1 that is causing the difference in L2.

This is familiar from the psychological case. Sometimes a change in a psychological state is evidently due to a change in a physiological state. For example an aspirin may make a headache go away. This is like Locke's case in which pounding an almond changes its colour. But sometimes a change in a psychological state seems to be due to a change in another psychological state, as when a piece of good news makes my headache go away. This is like the case in which we manipulate the colour of an object by using paint or ink or dye. The trouble is that sometimes we are unsure which kind of case we are dealing with. Suppose I find that when I am worried I have trouble sleeping. Is this because the worry is causing insomnia, or is it rather that there is some neural arousal that is constituting my worrying, and that neural arousal is keeping me awake? To what principles should we be appealing in addressing this problem?

The manipulationist account I sketched earlier provides no immediate way of answering this question. The approach simply assumes that we have already identified a suitable set of independent variables, and that the notion of an intervention is so carefully defined that if there is a change in the value of one variable when there is an intervention on another, that can only reflect a causal connection between the two variables. It does not immediately provide a way of addressing the question which of two non-independent variables, H1 and L1, should be thought of as causing H2.

There is, however, a natural way of developing the manipulationist approach so that this question is addressed. The manipulationist's idea is that to say that one variable is a cause of another is to say that manipulating the first variable is a way of manipulating the second. So far, the only aspect of that idea we have used is that there are some cases in which making a difference to the first variable will make a difference to the second

variable. But there is more to the ordinary notion of manipulation than that. Typically, we think of varying one variable as a way of making a systematic difference to the second. That is, there is such a thing as setting the first variable to one value then another and having there be a correlative change in the values of the second variable; as when we systematically affect how much water is coming out of a tap by turning the tap more or less. How much water comes out varies systematically with how far the tap is turned.

Here I am thinking of manipulation as a specifically human intentional action: you intentionally control one variable by affecting another. But in this intentional action you are exploiting a prior relation that is defined quite independently of the possibility of intentional action: for example, the independent relation between the degree of rotation of the tap and the amount of water coming out. It is this second, independent, systematic relation between variables that I am saying we should look at when considering the question at what level we find causal relations.

In the clear cases in which mental state is being manipulated by some physical variable, as when aspirin cures a headache, there is systematic correlation between a physical variable and the content of the consequent psychological state. That is, the more aspirin you take, within limits of course, the greater the pain relief offered. And similarly for any other case in which it is clearly the manipulation of a physical variable that is causing the change in psychological state. Or again, the more the almond is pounded, the dirtier the colour produced.

Still, I do not think it would be right to suggest that there must always be a systematic relation between the values of two variables for one to be a cause of another. Surely we can envisage cases – surely there are cases – in which there is just one value of

the first variable that has a single, quite specific effect on the value of the second, and nothing more systematic than that to be found. So it does not seem promising to look for an absolute condition on what it takes for one variable to be a cause of another, in terms of systematicity. What I want to propose is rather a relative criterion. In the case in which we have both a high-level variable and a type of low-level variable on whose values the high-level variable supervenes, we should describe the causation as high-level causation if we can find more systematicity in the ways in which H1 can be manipulated to change H2, than we can find in the ways in which L1 can be manipulated to change H2.

The mere fact that the mental is entirely constituted by the physical does not of itself mean that there will in general be any systematic variation of mental variables as a result of change in physical variables. Continuous variation in the physical variable B might be accompanied by apparently random changes in which psychological state, if any, ensued. In contrast, systematic changes in the psychological content of the intervention might be accompanied by systematic changes in the psychological content of the upshot. In that case, there is certainly a contrast with the case of the aspirin, and we can mark the difference by saying that here it is the mental variable whose manipulation is responsible for the variation in the subsequent psychological state. I think that approach simply reflects scientific practice. Consider again the case of worry and insomnia. Should we say that the worry is causing the insomnia, or should we say that the neural arousal is causing the insomnia? If cognitive interventions on the worry have a systematic effect on the insomnia we will say that the worry is the cause; if physiological effects on the level of arousal have a systematic effect on the insomnia we will say that

the arousal is the cause. It may also be that we will not have to choose: it seems entirely possible that insomnia should vary systematically with both worry and some purely physiological measure of arousal.

Finally, notice that this picture of high-level causation makes no appeal to the notion of a mechanism; and it allows for the possibility of effects being produced by combinations of high-level and low-level variables. There may be cases of colour change which illustrate this kind of possibility; but there are certainly many possible cases to be found in psychiatry. Consider a recent finding, that an early episode of humiliation is one of the causes of later depression. Not everyone is affected in this way by humiliation; some are resilient in the face of adversity. It may be that what constitutes resilience here is a normally functioning serotonin system; it may be that what constitutes vulnerability is an eccentricity in the serotonin system. And it may be that this physiological variable interacts with the psychological variable – humiliation – to produce depression, and that there is no further story to be told about any mechanism linking the physiological and the psychological variable. There may be no systematic account to be given of the physiological realisation of humiliation. Confronted with this possibility, it is natural to protest that there must be a mechanism linking the variables. But it is not easy to discover any justification for the idea that there must be a mechanism here, or even any articulated account of what we mean by 'mechanism'. Newton's account of gravity found an interaction between variables of attraction, mass and distance, but no mechanism to link them. And it was natural to protest that there somehow had to be a mechanism. But it is not easy to explain why we should think that mechanisms will always be found, or, indeed, what they are.

REFERENCES

Dummett, Michael. 1973. *Frege: Philosophy of Language*. Oxford: Blackwell.

Geschwind, Norman and M. Fusillo. 1997. 'Color-Naming Deficits in Association with Alexia'. In Alex Byrne and David Hilbert (eds.), *Readings on Color, Volume. 2: The Science of Color*. Cambridge, Mass.: MIT Press.

Gopnik, Alison, C. Glymour, D.M. Sobel, L.E. Schulz, T. Kushnir and D. Danks. In press. 'A Theory of Causal Learning in Children: Causal Maps and Bayes Nets.' *Psychological Review*.

Locke, John. 1975. *An Essay Concerning Human Understanding*, edited by P.H. Nidditch. Oxford: Oxford University Press.

Spirtes, P., C. Glymour and R. Scheines. 1993. *Causation, Prediction and Search*. New York: Springer-Verlag.

Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

Woodward, James and C. Hitchcock. 2003. 'Explanatory Generaliizations, Part 1: A Counterfactual Account'. *Nous* 37, 1-24.