

## **MANIPULATING COLOUR: POUNDING AN ALMOND**

**BY JOHN CAMPBELL**

It seems a compelling idea that experience of colour plays some role in our having concepts of the various colours, but in trying to explain the role experience plays the first thing we have to describe is what sort of colour experience matters here. I will argue that the kind of experience that matters is conscious attention to the colours of objects as an aspect of them on which direct intervention is selectively possible. As I will explain this idea, it is a matter of being able to use experience to inform linguistic or conceptual thought about what would happen were there to be various interventions on an object.

Against this background, I will review Locke's fundamental argument that since we can change the colour of an almond by pounding it, there must be an error embodied in our ordinary concepts of colour: there is no such thing as intervening directly on the colour of an object. The analysis I present brings out the force of Locke's argument. But I will propose a vindication of our commonsense conception of colour as an aspect of objects on which direct intervention is selectively possible.

### **1. 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 demanded 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 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.

An analogy might be helpful. Our ordinary visual world is full of shadows as well as highlights. And these shadows and highlights are very important to us in allowing us to see how objects are oriented with respect to us, providing some indications as to the character of the illuminant, and so on. But you could accept that is so while still pointing out that you might go through your whole life without ever paying attention explicitly to the shadows of the things around you. You could be such an inattentive person; inattentive just to the shadows, that is, though they do help you, did you but know

it, in discriminating and perceiving the objects around you. Just so, someone could use the colours, did she but know it, to allow her to identify objects, but never have attended to those colours. We might find this hard to imagine. We might find it hard to imagine because we find it hard to imagine not attending to so salient an aspect of our environment. That idea is consistent with the main point I want to make, which is that it is one thing to have full colour vision and be experiencing a coloured environment, and it is a further matter whether you attend to the colours in your environment.

We can draw a distinction, then, between colour as an object-defining characteristic, in the sense in which it is only the colours of the blobs in the display that define the figure 5, and colour as a characteristic to which the subject attends. Colour can be functioning as an object-defining characteristic even though the subject is not yet able to attend to colour. Notice, though, that if colour is to be functioning as an object-defining characteristic for a perceiver, the colours of the things seen must be showing up in the visual experience of the subject. If the colours of the blobs were not showing up in the experience of the subject, there would be nothing in the experience of the array itself which would differentiate the figure 5 from its background. Suppose a subject can see the figure 5 in the display I described. The subject has, we can suppose, ordinary colour vision, but as yet no ability to attend specifically to the colours of things. It seems quite evident that this subject need not have realised that there is such a characteristic as colour, despite the fact that the content of her vision includes experience of the colours. The natural formulation, in the light of what I have said so far, is that 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 she sees.

There are, though, many tasks that involve attention to colours which do not seem to involve an understanding of colour concepts. Suppose we have a subject who performs the following tasks. When given two rows of coloured paper, she can match each paper in one row to the same-coloured paper in the other row. Or again, when she is given a pile of chips of two slightly different shades of green, she sorts them successfully. She 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, she can correctly match the colours to the objects, for example, the yellow crayon to the banana. In performing these tasks she is plainly attending to the perceived colours of objects. Let us suppose that she also passes the tests I mentioned earlier: she can use colour as an object-defining property. She 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. Yet this attention merely for the purposes of matching colour samples, together with the use of colour as an object-defining characteristic, does not seem to be enough for knowledge of what the colours are.

It is not, though, as if what is missing is a battery of purely verbal skills. 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.

There was not, however, a 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. How are we to characterise the kind of liaison between colour names and colour vision that is required for grasp of colour concepts?

So much for an initial statement of the problem I aim to address. I think that the best way to present the analysis I propose is to state immediately my response to the question. Then I will set out the general considerations about the notion of ‘grasp of a concept’ which seem to me to matter for evaluation of the response. Finally, I will look at a classical argument for error theories of colour.

## **2. Interventionism**

In this section I want to try to describe a particular kind of awareness of colour: awareness of colour as an aspect of an object on which intervention is selectively possible. I draw the notion of an ‘intervention’ I will be using from the literature on causal reasoning (for excellent philosophical discussions, see Woodward and Hitchcock 2003, Woodward 2003, and references therein). Suppose you have noticed that there is a correlation between smoking and cancer, or between the position of a speedometer

needle and the speed of a car, and you wonder in each case whether the first is causing the second. What exactly is it that you are asking? The interventionist proposal is: you are asking whether, were there to be an intervention on the smoking, there would be a difference in the level of cancer, or, were there to be an intervention on the position of the speedometer, there would be a difference in the speed of the car. This is an intuitively appealing analysis, but evidently it depends on it being possible to explain just what is meant by an ‘intervention’. Not just any way of affecting the target variable will do. If your only way of affecting the position of the speedometer pointer is by affecting the speed of the car, then you will, trivially, find that the position of the speedometer continues to be correlated with the speed of the car under such ‘interventions’. What we want to be asking is whether, if some external force comes from outside the system and changes the position of the pointer on the speedometer – for example, if someone physically grasps and moves the pointer – there would be a corresponding change in the speed of the car. If there would be such a change, that constitutes the existence of a causal link between the position of the pointer and the speed of the car. Of course, there is no such link in this case, though there is in the case of smoking and cancer: external interventions on the level of smoking are correlated with changes in the level of cancer.

Just to spell out the notion a bit more fully. An ‘intervention’ on one variable  $X$  with respect to another,  $Y$ , should be a way of affecting the value of  $X$ , and ideally it should take full control of the value of  $X$  (when moving the speedometer needle you want to have total control over the position of the needle, and to have suspended the usual control of the position of the needle by the speed of the car). And the way in which you affect the value of  $X$  should affect only the value of  $X$ , and not have any impact on the

value of Y otherwise than by affecting the value of X. Moreover, you do not want to be any bias – you do not want to find that intervening on X is correlated with cases in which, as it happens, there was going to be a change in the value of Y anyway. Suppose we are armed with this notion of an intervention. And suppose it's true that, were there to be an intervention on X, there would be a change in the value of Y. On the interventionist analysis, that constitutes the existence of a causal relation between X and Y. (Here I follow Woodward and Hitchcock 2003, Woodward 2003, and the interventionist tradition within which they are working.) I want now to focus on how this notion of an intervention might illuminate our ordinary understanding of colour concepts, our knowledge of what the colours are.

It is often observed that the colours of objects have predictive value. The particular colours of various foods are predictive of their nutritional value. The exact colours of particular people and plants are good predictors of their health. And so on. Even though the correlations are typically specific to particular types of context, they have some generality. It is important, though, in considering such cases, to distinguish between colour as a symptom and colour as a cause.

I suggest that our ordinary understanding that colour is in these cases a symptom rather than a cause is provided by our grasp of what will happen under interventions. To say that colour is merely a symptom of nutritional value or of health is to say that nutritional value and health cannot be manipulated by manipulating the colours of objects. Consider the contrast between colour on the one hand, and shape or size on the other. Being able to attend specifically to such dimensions as the size and shape of a seen object means that one grasps the implications for other variables of interventions on the

size and shape of the seen object. You know what would happen if there were an intervention from outside to affect the size and shape of the thing. So, for example, you might squash and compress an envelope to get it through a letterbox. You know how specific sizes and shapes of envelope are correlated with the possibility of getting the thing through a letterbox. But you do not just have knowledge of correlations here. You have something more: knowledge of what would happen were there to be interventions on the size and shape of the envelope. Contrast the case of knowing, say, the correlation between the redness of a tomato and how ripe it is. You may use that information when choosing which tomato to eat. But you have some grasp of causal role here too, although of a different kind. Even if you are able to intervene on the colour of the object – say, by painting it – it would simply not occur to most of us to think that you could affect the ripeness of the tomato by manipulating the colour of the thing. This contrast between the kinds of manipulations we would ordinarily try to perform on shape, size and colour displays something of our ordinary grasp of the causal roles of these properties.

To sum up, when we think about interventions on the colour of an object, we find some quite special characteristics of colour. In the case of shape, there are many purposes to which we can put manipulation of shape. 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 causal significance. 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. You can't, though, 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 water was deep 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.

We usually take it that perception of an object as having a 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. What this distinction comes to, I think, is that we ordinarily take it that there is a difference between changing the colour of the object itself, and merely changing the way it looks to an observer by manipulating the conditions of perception. We ordinarily experience the colours of objects as dimensions of them on which intervention is selectively possible. There are indeed cases, such as the colour of a star in the night sky, where we have no idea what it would be to intervene to affect the colour of the thing itself, as opposed to merely affecting our perceptions of colour. And these, of course, are cases in which we have no idea what it would mean to talk about the colour the object itself has, as opposed to the appearance it presents. In contrast, we would usually have no trouble in distinguishing between making a door look white by shining a bright light on it, and making the door look white by painting it white; in the latter case, but not the former, there has been a change in the colour of the thing itself. We experience colour as an aspect of an object on which intervention is selectively

possible. That is, there are external causes which can change that aspect of the object in particular. Most strikingly, there are the paints, pigments, inks and dyes that have reliable specific effects on the colours of things. There are also causes that affect many manifest aspects of an object simultaneously, as when fire scorches an object, affecting its colour but also perhaps melting it. The contrast here between merely changing perceptions of the object, as opposed to changing the colour of the object itself, shows up in how the change affects perceptions of the object in different types of context. If I just shine a light on the door, that will have no implications for how would look in ordinary sunlight. If I paint the door, in contrast, that will have implications for how it would look in sunlight; and in other types of context too.

My question has been how to characterise the type of liaison between colour experience and verbal or conceptual thought that is demanded for grasp of colour concepts. We saw that it isn't enough merely that you have colour experience; you must be able to attend specifically to the colours of objects. But then, what type of attention is needed? I want now to argue that the kind of visual attention that is needed is attention to colour as an aspect of objects on which selective intervention is possible.

### **3. Concepts: Truth-Conditions vs. Reasons**

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 (Dummett 1973). 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.

The account I am proposing of the role of experience in an understanding of colour concepts appeals to this classical conception of what it is to understand a predicate. The proposal is that the role of colour experience is to provide knowledge of various mappings from objects to truth values. Colour experience can be seen to play that role when we conceive of colour experience as a matter of attention to colour as an aspect of objects on which selective intervention is possible.

There are two different levels at which you might have the conception of colour as an aspect of objects on which selective intervention is possible. It might be an entirely practical matter, of the skills that you have in interacting with your surroundings. You might have a capacity to affect the colours of objects by whatever means – paints or inks, for example – and you might in practice be able to recognise the upshot of those interventions for your own experiences, and those of other people. There might in this be some implicit recognition that your own actions are of the same general types as those of other people: for instance, you might be able to imitate the interventions of other people, or to recognise when you are being imitated yourself. You could have this practical grasp of colour as an aspect of objects on which selective intervention is possible without having any explicit conception of experience at all; it may never have occurred to you explicitly that you and others have experiences of the world, you just are in practice able to affect what those experiences will be by manipulating the colours of objects. Someone

who has this capacity has evidently gone far beyond a subject who is capable merely of matching colour samples, or using colour as an object-defining property.

When I speak of ‘intervention being possible’ I am not talking about the possibility of specifically human action; it is the general notion of something external making a change in the colour of the object, of which human action is one example. And it is this modal fact that we exploit when in practice we do manipulate colours. So we can contrast the purely practical understanding I just described, of colour as an aspect of objects on which selective intervention is possible, with a theoretical grasp of colour as an aspect of objects on which selective intervention is possible. You can have a theoretical understanding of those modal facts which does not simply consist in the capacity to exploit them in manipulating the colours of objects. Attending to colour as an aspect of the object on which selective intervention is possible can be a purely practical matter of the range of interactions with the object of which you are capable. But attention to colour can also have to do with your theoretical understanding of the modal facts that you exploit when interacting with the object. This kind of understanding has to do with the way you would use the names of colours in saying what would happen in various counterfactual situations, such as those in which there are interventions on the colour of the object. And it is conscious attention to the colours of things, informed by this general theoretical understanding, that, I suggest, constitutes grasp of the ordinary colour concepts. It is when you have reached this point that you have a grasp of what it is for it to be true that an object has a particular colour.

Our commonsense picture of colour is that the observed colour of an object is the very property on which we intervene, when we act to change the colour of the object.

When we ink over or paint or dye an object, we take ourselves to be acting directly on the very property of the object that we observe; we do not assume that the ink or paint or dye operate directly on some quite hidden aspect of the object, and only consequently affect the observed colour of the thing. It is in this sense that we do not ordinarily suppose colour to be merely a power that objects have to produce experiences in us. Were colours mere powers, it would not be possible to affect them directly; you could affect them only by affecting their bases. But we do assume that we can affect the observed colours of objects directly. And even if we cannot in practice do this, because of the limitations of the technologies available to us, we take it that when we observe the colour of any object, we are observing an aspect of it on which direct intervention is in principle possible. That is, we take it that we are observing a categorical property of the object.

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. I will call this the 'pure reasons-based approach'. The classical approach I have just recommended did not discuss the notion of experience providing reasons for belief at all. The pure reasons-based approach did not discuss the notion of grasp of truth-condition at all. Obviously a variety of mixed views are possible; but let me pursue the pure reasons-based approach for a moment.

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 pure 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 the pure 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 pure reasons-based approach might acknowledge 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 the pure 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.

Many philosophers – the classical sources are Galileo and Locke – have 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 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.

The pure reasons-based account, as I have developed it, makes error theories of colour impossible. The account is in effect arguing that we could not have the conception of colour as a categorical characteristic which, the error theorist says, science has shown to be mistaken. For, the pure 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, even if we acknowledge that it has to be supplemented with an account of the role of experience in providing reasons for colour judgements. 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. This is the conception of colour as a categorical property, which can be specifically manipulated.

#### 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 concept of colour as categorical. On this account, knowledge of the colours is provided by conscious attention to colour as an aspect of the object on which direct intervention is selectively possible. 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.

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 of 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. The variables are not independent. 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.

The issue depends on which set of variables it is right to use in describing the phenomena here. If the choice of a set of variables is arbitrary, then the issue has no substance. But in general the choice of a variable set does not seem to be arbitrary; we would usually think of it as one of the most difficult matters to address in finding how to characterise the causal functioning of a system.

I want to make a proposal about the general type of consideration we ought to be appealing to here, a proposal which is, I think, in line with the general spirit of an interventionist approach to causation. Here is a simple example to illustrate the idea.

Suppose you are asked at what level you should characterise the causal functioning of a radio. You want to know whether the true causal structure is to be found at the level of a quantum-mechanical description of the whole set-up. So macroscopic matters such as the position of the volume control and whether the tuner has been set to a particular station are not themselves part of the causal structure; they are merely epiphenomena dependent on the underlying quantum-mechanical causal structure. Now it seems to me that an interventionist approach to causation has the materials to motivate the idea that we do find causal structure at the level of the macroscopic variables such as the position of the volume control. An interventionist approach is not anthropocentric; it does not aim to characterise causation in terms of what humans can do. But it does aim to describe those objective features of our world that we exploit when we manipulate our surroundings. Now the point about the relation between, say, the position of the volume control and the loudness of the sound from the radio is this. Under interventions on the position of the volume control, there is a correlation between each particular position of the volume control and each particular level of loudness. But there is more to it than that. There is a certain systematicity in this correlation under interventions: the level of loudness varies with the position of the control. Moreover, there is a very large statistical effect here. And finally, the effect is specific: the position of the volume control selectively affects the loudness of the sound, it is not nearly so strongly correlated with any other outcome. These are objective features of the set-up, though they are of course the features we exploit in manipulating the controls. And I propose that if we can find a level of description, a collection of variables to use in characterising a system, that has these

features, then that constitutes the correctness of saying that the causal functioning of the system can be characterised in terms of those variables.

So one response to Locke's argument is simply to acknowledge the correctness of his 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. The point about pounding is that it is an effective systematic control specifically for such variables as motion. It is not an effective systematic control specifically of colour. So if all interventions on colour were of this type, then we might accept that colour is not the right variable in terms of which to characterise the changes that mediate between an intervention on the object and subsequent changes in the colour experiences of those who see the thing.

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 specifically 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. But the whole point of these substances is that they have large, systematic effects specifically on the colours of a variety of objects. It is not an appeal to magic to propose that we employ a set of variables characterising the interventions here under which 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 of itself does not show that the only causality here was at the level of the supervenience base. Our knowledge of the ordinary causes and effects of colour change, of the workings of paints and pigments and inks and dyes, is part of a common-sense 'colour theory', describing the large systematic upshots of interventions, which allows us to regard to colours as categorical properties of objects, mediating between intervention on the object and the consequent changes specifically in the colour experiences of those who see the thing.

The criteria I am setting out here were foreshadowed in the medical statistician Hill's classic article giving criteria for the existence of a causal relation between an environmental hazard and a disease (Hill 1965). One criterion he gave was the 'dose-response' criterion:

if the association is one which can reveal a biological gradient, or dose-response curve, then we should look most carefully for such evidence. For instance, the fact that the death rate from cancer of the lung rises linearly with the number of cigarettes smoked daily, adds a very great deal to the simpler evidence that cigarette smokers have a higher death rate than non-smokers.

(Hill 1965, p. 298)

Similarly, the case for the causal efficacy of a drug is enhanced if we find not merely that recovery from illness is correlated with administration of the drug, but that the degree of recovery from illness is correlated with the amount of the drug administered. A second criterion is the sheer size of the correlation between the hazard and the disease: that is,

the size of the ratio of the rate of contraction of the disease among those exposed to the hazard to the rate of contraction of the disease among those not exposed to the hazard. And the third of his criteria I want to mention here is the specificity of the correlation between the hazard and the disease.

Hill is explicit that he is not attempting to give an analysis of what causation is; these criteria, and the others that he gives, are intended as guides to when the practitioner has the right to conclude that there is not merely an association but a causal relation between two variables. It is because these remarks are not aimed at the analysis of what causation is that they may seem to be of merely practical, rather than philosophical importance.

As a medical statistician, Hill is approaching the question of causation from a broadly interventionist standpoint:

with the aims of occupational, and almost synonymously preventive, medicine in mind the decisive question is whether the frequency of the undesirable event B will be influenced by a change in the environmental feature A.

(Hill 1965, p. 295)

But the criteria he proposes are not asking merely whether some change or other can be effected by one or another intervention. His criteria are asking: how good are these variables as systematic ways of bringing about large changes in specifically these selected outcome measures?

It is certainly possible to have causation without the possibility of this kind of systematic control over the upshot. But we can nonetheless view Hill's criteria as giving us the beginning of a constitutive account of when it is right to use one set of variables rather than another in characterising the causal functioning of a system. Suppose we find a correlation between an environmental hazard and a disease that meets Hill's criteria; for example, the correlation between smoking and lung cancer. It would be possible to insist that nonetheless, the true causal structure here is to be found at the level of quantum mechanics. The relations between smoking and lung cancer, you might say, are merely epiphenomenal. But here it seems to me that Hill's criteria do have constitutive force. Since they show that intervention on smoking can be regarded as one of the variables providing for effective systematic control over the degree of lung cancer prevailing, there is no further question as to whether the true causal structure is to be found at some more basic level. Indeed, once we move to the quantum mechanical level, we may well lose sight of any variables at all which would meet Hill's criteria for causation of lung cancer.

### **5. Causation without Mechanisms: The Psychological Case**

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-level 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 interventionist terms, the problem arises because when you intervene on a high-level state you cannot but be intervening on 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 (cf., e.g., Kim 1997).

How does the approach I have been sketching bear on causation in psychology? 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 interventionist account I sketched in §2 of itself provides no immediate way of answering this question. That 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, 'worry' or 'neural arousal', should be thought of as causing wakefulness.

I have, though, been proposing that there is a natural way of developing the interventionist approach so that this question is addressed. To characterise the causal functioning of a system, we have to find a set of variables in terms of which we can characterise interventions on the system. And we should aim to find a set of variables that maximises the effectiveness, systematicity and specificity of the impacts of interventions on our outcome variables.

The mere fact that the mental is entirely constituted by the physical does not of itself mean that there will in general be any effective systematic variation specifically of mental 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 psychological state, if any, ensued. In contrast, systematic changes in the psychological content of the intervention might be accompanied by large and systematic changes specifically in the psychological content of the upshot. In that case 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 large, systematic and specific effect on the insomnia we will say that the worry is the cause; if physiological interventions 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, I want to end with one further remark on Locke's challenge. I have looked at just one element in the challenge: the problem of finding the right variable set to use. The other element is his appeal to a mechanistic view of causation. In effect, he is arguing, causation just is the transmission of motion by impulse. If we are to regard the pounding of an almond as causing change in colour, we have to suppose that there is nothing to the change in colour other than a change in certain motions. The colour we observe is not in fact the property we are acting on in an intervention. We would not now regard this form of mechanism about causation as tenable; there are plainly many causal interactions that do not consist merely in the transmission of motion by impulse. But it is not difficult to find more recent versions of mechanism about causation in terms of which it is easy to reformulate Locke's challenge. We can, for example, appeal to the proposal put forward by Dowe 2000, that causal interaction involves the exchange of conserved quantities. Since colour is not a conserved quantity (that is, a property subject to a

conservation law – there is no law of the conservation of clear whiteness, for instance, unlike the situation with mass-energy, linear momentum or charge) it cannot figure in causal interactions. Therefore, the argument runs, we cannot be manipulating colour.

Notice, though, that the picture of high-level causation I have sketched 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 (Kendler et. al. 2003). 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, humiliation and serotonin imbalance. But what mechanism could this be? It could not be a purely cognitive mechanism, because the serotonin imbalance is a biological phenomenon. We can make nothing of the idea of a ‘mechanism’ linking the experience of humiliation and this biological phenomenon – unless, of course, we think that we can give a reductive biological account of the

experience of humiliation. And perhaps we can make something of the notion of a straightforwardly biological mechanism. Now, of course, biological reductionism may turn out to be correct. But in the present state of our knowledge, it is reckless to say that it must be correct. We could still have knowledge of the existence of a causal relation between humiliation, serotonin imbalance and later depression. The idea that the mere existence of a causal relation means there ‘must’ be a mechanism implies that we cannot recognise the causal relation without the reckless commitment to reductionism. We should rather let go of the apparently innocuous claim that there ‘must’ be a mechanism. We do not need any such commitment to acknowledge the truth of counterfactuals about what would happen to a system under interventions, where the variables characterising the system are identified using the criteria I have indicated. And that is all we need to talk of causation.

## ACKNOWLEDGEMENTS

I have benefited from discussion of early drafts at NYU, Oxford, the University of California at Santa Brabara, the University of Southern California and the Center for Advanced Study at Stanford. I am particularly grateful for comments by Kevin Falvey when I presented this material at UCSB, and to Jerry Fodor and Christopher Peacocke for comments at their NYU seminar. Victor Caston also gave me a helpful set of comments. Thanks also to Alison Gopnik, Thomas Richardson and Ken Kendler, and to Christopher Hitchcock and James Woodward.

## REFERENCES

Dowe, Phil. 2000. *Physical Causation*. Cambridge: Cambridge University Press.

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.

Hill, Austin Bradford. 1965. 'The Environment and Disease: Association or Causation?'. *Proceedings of the Royal Society of Medicine* 58, 295-300.

Kendler, Kenneth S., John M. Hettema, Frank Butera, Charles O. Gardner and Carol A. Prescott. 2003. 'Life Event Dimensions of Loss, Humiliation, Entrapment, and Danger in the Prediction of Onsets of Major Depression and Generalized Anxiety'. *Arch Gen Psychiatry* 60, 789-796.

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

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

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