AN INTERVENTIONIST APPROACH TO CAUSATION IN PSYCHOLOGY

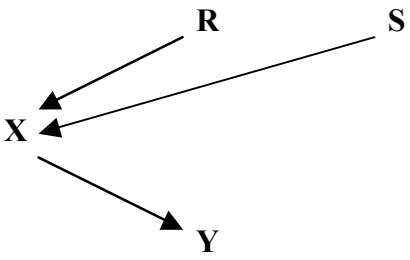
BY JOHN CAMPBELL

My project in this paper is to extend the interventionist analysis of causation to give an account of causation in psychology. Many aspects of empirical investigation into psychological causation fit straightforwardly into the interventionist framework. I address three problems. First, the problem of explaining what it is for a causal relation to be properly psychological rather than merely biological. Second, the problem of rational causation: how it is that reasons can be causes. Finally, I look at the implications of an interventionist analysis for the idea that an inquiry into psychological causes must be an inquiry into causal mechanisms. I begin by setting out the main ideas of the interventionist approach.

1. Interventionism

Interventionism is the view that for X to be a cause of Y is for intervening on X to be a way of intervening on Y (cf. Woodward and Hitchcock 2003, Woodward 2003, Pearl 2000, Spirtes, Glymour and Scheines 1992). The interventionist approach can be vividly expressed by means of causal graphs, which use arrows to depict causal relations between variables. These arrows may represent positive or inhibiting causal relations. Suppose we consider a causal relation between variables X and Y. Suppose for example that X

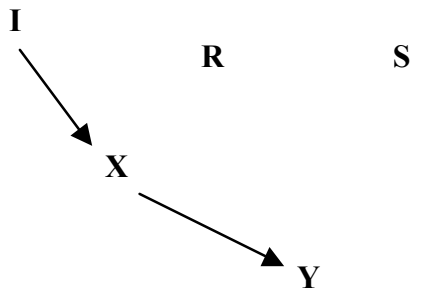
represents the level of a drug in someone's blood and that Y represents whether and how well the subject recovers from an illness. Suppose further that the body endogenously produces the drug, in varying quantities in different people. There will be some biological factor responsible for the level of endogenous production of the drug in someone's body; suppose we express this by variable R. And suppose that the drug is also spontaneously ingested by people as part of their ordinary diet, in varying amounts by different people; suppose we summarize the factors responsible for spontaneous ingestion of the drug in ordinary diet by variable S. Then we can represent the hypothesis that the level of the drug is a cause of degree of recovery from the illness as follows:



The arrows show variables R and S causally affecting X, and X causally affecting Y. The objective of an interventionist analysis is to explain what it is for X to be causally affecting Y. The intuitive idea is that for X to cause Y is for intervening on X to be a way of intervening on Y (intervening on the level of drug will be a way of intervening on degree of recovery from the illness).

Following Woodward and Hitchcock 2003, we can exhibit an intervention on X in terms of a variable I which acts on X. (For instance, we might think of an external agent giving people various amounts of the drug, and observers keeping track of the subsequent

degrees of recovery of people from the illness.) The idea then is that there are at any rate some circumstances in which were there to be an intervention on X, there would be a difference in the value of Y.



There is a possibility that R and S might be common causes of both X and Y. In that case variations in X will be correlated with variations in Y, but that may not be because X causes Y. (So, for example, we have to keep in mind the possibility that the factors which cause endogenous production of the drug, or which lead a person to ingest a lot of it, might each be a common cause of both the level of drug in a person's body and the degree of recovery from the illness. In that case, we will find that there is indeed a correlation between degree of recovery and level of drug in the body, but that will not constitute a causal relation between the level of drug and the degree of recovery. So we should want an intervention on X to suspend the influence of these other factors on the level of drug in the blood.) In general, then, the intervention variable I should take over control of the value of X, removing it from the influence of R and S. To use Pearl's term, the intervention should be 'surgical', breaking the arrows from R and S to X. Given that condition on the intervention variable I, then we can say that for X to cause Y is for it to

be the case that there is a correlation between X and Y under potential interventions on X.

There are further conditions to be met. We have to exclude the possibility that the intervention I on X also affects Y directly. (For example, administering the drug should not have a placebo effect.) So we should stipulate that an intervention variable for X with respect to Y must not affect Y otherwise than by affecting X. We should require that there is no bias in which interventions are carried out; that is, that there should be no correlation between intervention and recovery (for instance, we should not be administering the drug only to those who are going to recover anyway). And finally, we should have a requirement of 'causal sufficiency' on the variables we have explicitly represented; in particular there should be no unrepresented variables which are common causes of pairs of variables which we do have explicitly represented, so that spurious correlations can be generated.

With these stipulations in place, though, we can define what it is for X to cause Y by saying that were there to be an intervention on X, there would in some cases be a difference in the value of Y. Or, equivalently, we can say that for X to cause Y is for X and Y to be correlated under potential interventions on X. This is not a reductive definition of causation. On the contrary, it makes free use of causal notions in defining the idea of an intervention, and in explaining what it is for a set of variables to be 'causally sufficient'. Nonetheless, the definition I have just given does not appeal to the idea of a causal relation specifically between X and Y. It has therefore some claim to provide a non-reductive illumination of the notion, by locating it in a broader framework of causal notions.

In my remarks in this section I have followed very closely the approach to causation developed by Woodward and Hitchcock, building on the earlier work of Pearl and Spirtes, Glymour and Scheines; any originality so far is accidental. Notice that the approach presupposes a certain modularity in the system of variables in question. It presupposes that interventions on the system can in principle leave undisturbed the causal relations among particular variables. That is, an intervention can selectively disturb certain causal relations – those involving the usual causes of the target variable X – while leaving others intact – in particular, the causal relation between the target variable X and the outcome variable Y (cf. Woodward and Hausman 2003).

2. Control Variables

I want now to ask whether this approach can be used to illuminate causation in psychology. On the face of it, there should be no special problem here. Consider any psychological variable M1, and the hypothesis that M1 is a cause of some other psychological variable M2. So, for example, consider the hypothesis that worry is a cause of insomnia. For worry to be a cause of insomnia is, on this approach, for it to be the case that were there to be an intervention on worry, there would be a difference in the level of insomnia. The trouble with this, though, is that any intervention on worry is also going to be an intervention on some underlying set of biological variables. You cannot affect worry without affecting the underlying biology. So how we describe the situation? Is it that the worry is causing the insomnia – that intervening on the worry is correlated

with a difference in the insomnia? Or is it rather that there is a biological variable underlying the worry, and it is causally related to a biological variable underlying the insomnia? In that case, the situation is better described by saying that intervention on the first biological variable is correlated with the second biological variables. The psychological variables, in that case, are epiphenomenal on the underlying biological causation.

Think how you would characterize the relation between the positions of the controls on a radio and the output of the radio, such as the volume of the sound or the radio station being heard. All that goes on here does indeed supervene on a microphysical reality. But we would ordinarily have no hesitation in saying that someone turning the controls is making a difference to the output. Why does it seem so evident here that the position of the dials is causing the output, and that we are not here dealing merely with epiphenomena?

I think we can get at this by recalling a famous set of criteria proposed in 1965 by the epidemiologist Austin Bradford Hill, to determine whether particular environmental hazards were causes of particular diseases, or merely correlated with them. Central among Hill's criteria are: (a) the existence of a dose-response effect. Most simply, this demands that there be an identifiable relationship between the value of the input variable and the correlated output variable. To demonstrate that smoking is a cause of cancer, for example, one critical piece is the datum that the amount one smokes is correlated with the probability of death from cancer. (b) It enhances the case for saying that smoking causes cancer if there is a large effect of smoking on cancer. And finally, (c) it enhances the case for saying that smoking is a cause of cancer if smoking is correlated specifically

with cancer, rather than any other outcome. I can sum this up by saying that the case for saying that smoking causes cancer is a case for saying that smoking is a *control variable* for cancer. Here I am using 'control' in the sense in which the buttons on a radio are controls. There is a large, specific and systematic correlation between the volume coming out of the radio as you twist volume knob. Just so, under interventions on the level of smoking, there are large, specific and systematic effects on cancer.

I am proposing that we should use this notion of a 'control variable' to identify the level at which we find the causally significant variables in a complex system. I think there is no question but that in the case of the radio, the positions of the various buttons and knobs are 'control variables' in this sense, and that this is why it seems so evident that making a difference to the controls of the radio is making a difference to the upshot; we are not dealing here with epiphenomena. For the case of smoking, consider how you would react to a spokesman for the tobacco industry who argued that smoking is not a cause of cancer; that smoking and cancer are both merely epiphenomenal upon an underlying microphysical reality at which the true causal relations are to be found. The natural point to make in reply is that smoking is a control variable for cancer; interventions on smoking have large, specific and systematic correlations with cancer. That is the case for saying that the causal relations between smoking and cancer are to be found at the macrophysical level.

Of course, it will be a matter of degree whether one variable functions as a control variable for another, and there will be a certain relativity to context. But that is how it is with causal ascription generally, anyhow. Hill did not explicitly formulate his criteria as criteria for choice of variables to use in characterizing the data; and what I have said here

by no means exhausts his points. But the force of the idea ,that we find the right level at which to characterize causal relations by looking for the level of control variables, seems undeniable.

One way to see the force of that idea is to look again at the background picture of an interventionist approach to causation. An interventionist approach sees the interest or point of our notion of cause as having to do with our own manipulations of our environment. It is not that the notion of cause is being explained in terms of agency; it is, rather, that to characterize causal relations is to characterize the aspects of the world that we exploit when we manipulate it. If you think of causation in this way, then it seems evident that control variables will be of great importance in describing causation. For in manipulating the world we want, so far as possible, to be intervening on variables that are correlated with large, specific and systematic upshots. We want to be intervening on control variables in our actions. In these terms, then, the case for saying that worry is a cause of insomnia is that worry is a control variable for insomnia. What is it to say that worry causes insomnia, and that the two are not merely epiphenomena? It is to say that interventions on worry are correlated with large, specific and systematic variations in insomnia.

3. Causation by Reasons

Some difficult issues concern the application of the interventionist picture to what we might call ‘rational causation’: cases in which the causal explanation appeals to the

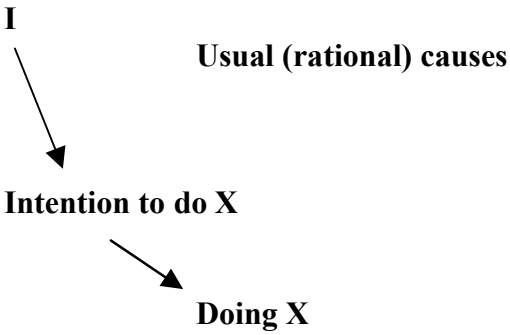
subject's possession of reasons. Suppose we consider, for instance, the hypothesis that the intention to do X causes doing X. Can we think of this in terms of whether there would be differences in whether X was performed if there were interventions on the intention to do X?

The really difficult thing here is to find the right characterization of a psychological intervention. What is it to intervene on whether someone has the intention to do X? We would naturally think of this in terms of providing someone with reasons to do X, or reasons not to do X. 'You think doing X will make you happy, but it won't', you might say as an opening move. And you might present further considerations in favor of your remark. You would be appealing to the rationality of the subject. The trouble with this is that it leaves intact the factors that are the usual causes of the someone's forming, or not forming, the intention to do something. For example, suppose that one of the usual causes of a person's intending to do X is that they think doing X will make them happy. If your 'intervention' takes the form of arguing about whether or not doing X will in fact make that person happy, then you have left in place one variable that is a usual cause of whether the person forms the intention to do X. This means that the intervention is not, in Pearl's term, 'surgical'. To use again the example of drug trial, suppose you are asking whether the level of drug in someone's body causes recovery from illness. If you manipulate the level of drug in that person's body by acting on the mechanism involved in the body's endogenous production of the drug, this does not constitute an 'intervention' in the sense I explained in the last section. Similarly, if an 'endogenous cause' of whether someone forms the intention to do X is whether the person believes that doing X will make them happy, a manipulation of whether the

person forms the intention that proceeds by manipulating whether the person believes that doing X will make them happy does not constitute an 'intervention' in the sense I explained.

The reason for insisting on a surgical intervention in the case of the drug trial was the problem of common causes: that the endogenous cause of the level of drug in the blood might also be directly causing recovery from illness, so that the level of drug in the blood actually played no role in causing recovery from illness despite being correlated with recovery. It is to rule out this scenario that we have to consider interventions that seize control from outside of the level of drug in the blood. Similarly, suppose we leave intact the 'endogenous causes' of formation of the intention to do X, such as the belief that doing X will make one happy. Then it is possible that the belief that doing X will make one happy causes both formation of the intention to do X and also directly causes performance of the action itself. In that case the intention to do X will be correlated with doing X even though the intention plays no role in causing the action. It is to rule out this scenario that we have to consider only 'surgical' interventions on the intention to do X, according to the interventionist picture as I have so far set it out.

What would it be to have a surgical intervention on someone's possession of an intention to do X? The intervention would have to come from outside and seize control of whether the subject had the intention, suspending the influence of the subject's usual reasons for forming an intention, such as whether the subject had reasons for forming the intention to do X. We can diagram the situation by means of a causal graph:



This is evidently quite an unusual situation. It does not happen very often, if it happens at all, that a person's rational autonomy is suspended and some alien force seizes control over whether that person has a particular intention. Still, even though it does not happen very often, it could still be that an interest in psychological causation is an interest in what would happen in such an unusual case. Similarly, you might say that an interest in causation in physics often deals with what would happen in various idealized conditions – in a complete vacuum, or on a frictionless plane, for example, even though such situations do not arise very often.

The real problem for the interventionist picture here is that it is not credible that our interest in psychological causation is an interest in what would happen under such idealized conditions of alien control. There are two aspects of our ordinary conception of the psychological life that have been removed in this scenario, and without them our psychological life would not be recognizable.

Notice first that ordinarily we have our intentions under continuous review. If you hit an obstacle in trying to execute your plan, you may review whether to sustain the intention in the light of all your background beliefs and objectives – just how important is this anyhow? – and how far you stick with an intention often depends on continuous review in the light of your other psychological states, your priorities and beliefs as to the

likelihood of success. If you could not do this kind of continuous monitoring, you would be said to be 'not responsible for your actions'. It is exactly this situation that we are envisaging, though, when we think in terms of 'surgical' intervention on possession of an intention.

Secondly, this scenario is one that would undermine our ordinary conception of the ownership of an intention. One element in our ordinary notion of the ownership of an intention is the idea that the long-standing objectives, interests, preferences and so on of that person were causally responsible for the formation of that particular intention. It is a reasonable description of the situation envisaged as surgical intervention here to say that someone else's intention has been thrust into the mind of the subject. Someone who seemed to find himself in that situation – someone who encountered in introspection an intention that seemed to have been the direct result of someone else's long-standing objectives, interests, preferences and so on – would experience this as 'thought insertion', the feeling that someone else's token thought has been pushed into your mind, one of the symptoms of schizophrenia.

There are many systems for which an approach in terms of surgical interventions seems appropriate. Suppose for example that our descendants come upon an archive of electrical machines, present-day radios, perhaps. And they want to find our just how the circuitry works. They are not concerned with the function of these devices. They just want to understand the electrical engineering involved. In this case an approach in terms of surgical interventions seems entirely apt. Even if it turns out not to be in practice possible to tear the systems apart into their modular constituents, still the objective is to find out what would happen in each constituent module were we to have a surgical

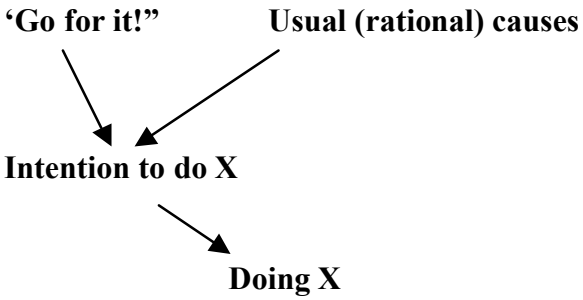
intervention that ripped out this piece of wiring from its context, and tampered with the input end to see what would happen at the output. We have understood the causal structure of the circuitry when we have answered all such questions. In the case of rational causation, in contrast, we have no such interest in ripping out individual pieces of circuitry from their context to see how they would behave in isolation. The attempt to do this would result in a system so different to the original that what happened in that context could not be said to have any significant implications for the functioning of the original intact system. This is a fundamental point about rational causation in psychology, which underpins some of the hesitation philosophers have felt in talking about mental causation at all.

4. Two Types of Intervention

I think that we can resolve this problem within a broadly interventionist framework, but that to do so we have to rethink our conception of an intervention; we have to move away from the focus on surgical interventions. We want to consider interventions that keep intact the rational autonomy of the subject, which means leaving in place the usual causes of the subject's psychological states and actions. But then what kind of thing are we looking for, to be a psychological intervention? Let me first give a couple of examples then give a more abstract statement of the general notion of intervention being presupposed.

Suppose that I am the passenger and you are driving as we come to a pool of water in the middle of the road. You stop to weigh up the situation. Should you drive on or should you back off? As you pause I say, 'Go for it!', and you put your foot on the accelerator. One possibility is that you have such admiration for my judgment and such concern to act as I would like that the mere fact of my making my remark of itself gives you a reason to form the intention to press on. However, that is not the most obvious or the natural analysis of the situation I have described. Perhaps you know that my judgment is in general questionable; perhaps you and I have just quarreled so that far from giving you a reason to form the intention to proceed, had you paused to reflect on the matter for a moment you would have found that my remark gives you good reason to swing round and go the other way. As it is, though, it is undeniable that my remark had the effect of making you form the intention to drive on, and that consequently you did drive on. In this case, my intervention affects the formation of your intention. But it does not do so by providing you with reasons for or against forming the intention. Rather, it directly affects the formation of your intention. I did manage to reach into your mind and affect the formation of your thought, otherwise than by giving reasons.

It is not, though, as if you had given over the reins of your mind to me. You remained an autonomous rational agent throughout. You could have resisted my remark; you may later regret that you did not do so. Had you had mustered reasons which struck you as compelling, one way or another, it could have been that my remark would have had little effect. The structure of the example can be given by the following causal graph:



The problem we encountered earlier with this kind of situation was this. We are attempting to explain the existence of a causal relation between the intention to act and the action as a matter of the intention and the action being correlated under interventions on the intention. But we have not yet excluded the possibility that the usual causes of the intention may also be direct causes of performance of the action. So even if the intention and the action are correlated under this kind of 'intervention' on the intention, it may be that this correlation is only a residue of the role of the usual causes of the intention in operating as common causes of both the intention and the action.

There is, though, another way in which we could think of interventions. Suppose we go back to the example of drug level and recovery from illness. Suppose we consider a range of actual or possible external administrations of the drug to individuals across a population. And suppose that when the drug is administered to an individual it is administered without the level of endogenous or spontaneous ingestion of the drug being taken into account; these factors are allowed to operate as usual. So this is not a surgical intervention. Nonetheless, we can look at the level of drug that is endogenously produced by the individual and at the level of drug that is spontaneously ingested by the individual. So for each combination of a particular level of endogenous production, and a particular level of spontaneous ingestion, we can consider what would be the outcome of

administering a particular level of the drug. And we can say: suppose that there is some combination of a particular level of endogenous production of the drug, and some level of spontaneous ingestion of the drug, such that were the external administration of the drug to be varied while those levels remained the same, there would be a difference in whether the subject recovered from illness. In that case, the level of drug in the blood is a cause of recovery from illness. In fact, on the way I propose of developing the interventionist account, this is what it is for the level of drug in the blood to be a cause of recovery from illness.

The difference between this formulation of interventionism and the analysis I reported in section 2 above emerges vividly when we consider cases, such as that of rational causation, for which modularity assumptions are not correct. We are not any longer considering whether the value of Y is independent of the value of X, when the value of X is set by surgical intervention. We are, rather, considering whether Y is independent of the intervention variable I given the usual causes of X. And the conditions that have to be met by the intervention variable I are just as before, except that we are no longer requiring that the influence of the usual causes of X should be suspended and that I should be the sole determinant of the value of X.

We can apply this picture to rational causation in psychology. We do not need to consider a scenario in which the rational autonomy of the agent is suspended, and some external factor seizes control of the agent's intentions. We can, rather, consider cases in which the usual causes of the agent's formation of intentions operate as usual, and look at whether external interventions which make a difference to whether the agent forms an intention, for some set set of values for the agent's other psychological states, would be

correlated with differences in whether the agent performs the action. Is intention a cause of action? My proposal is that this is the question whether interventions on intention are correlated with action, given the agent's other psychological states.

5. Psychological Causation Without Psychological Mechanisms

One of the most striking features of an interventionist approach to causation in psychology is that it makes no appeal to idea of mechanism. All that we are asking, when we ask whether X causes Y, is whether X is correlated with Y under interventions on X. Whether there is a 'mechanism' linking X and Y is a further question. Indeed, you could maintain an interventionist approach to causation while being skeptical about the very idea of a mechanism. What does it mean, to ask whether there is a mechanism linking X and Y? All that it comes to, you might say, is that we are asking whether we can find any causally significant variables mediating X and Y. Or perhaps, in some cases, we are asking merely that the link between X and Y should be explained in terms of one or another familiar pattern of explanation, for example biological explanation. But the very idea of a causal link does not demand that there should be intervening variables, or that assimilation to a favored paradigm should be available.

To see why this perspective matters, consider some recent findings in psychiatry. It has long been known that stressful live events such as bereavement or unemployment are good predictors of chronic depression. In a recent study of several thousand subjects, Kendler et. al. (2003) tried to determine which aspects of stressful life events might be

playing a causal role here. They found that the strongest correlations with later chronic depression were with humiliation rather than with loss; that other-initiated separation was a stronger predictor of chronic depression than bereavement, for example. To interpret the study as showing something about the causes of depression is to read it has having implications for what the upshot would be of clinical interventions: the implication is that under interventions to ameliorate the sense of humiliation there would be differences in the degree of chronic depression. In the sense I explained earlier, humiliation is a control variable, in the kind of non-surgical intervention I just described, for later depression.

Stressful life events, however, are not the only predictors of later depression: there are also biological factors that seem to be relevant. Kendler et. al. (in press) found that genetically acquired deficiencies in the serotonin transport system are correlated with later depression. Now, given the complexity of the phenomena, all such findings have to be regarded as provisional at the moment. In this paper I want finally to suggest, though, a simple reading of them, on which they provide a simple, illlustrative example of a quite general pattern emerging from current empirical work in psychology and psychiatry.

Although stressful live events predict depression, not everyone who is humiliated ends up with depression. People vary in how resilient they are. One reading of the serotonin data is that they reveal serotonin deficiencies to be the basis of a lack of resilience. On this reading, then, we have found two causal variables underlying later chronic depression: humiliation and serotonin deficiency. These are control variables for depression, let us suppose. And the relevant notion of intervention, let us suppose, of the kind I indicated above, where we consider psychological factors that affect the level of

humiliation directly, rather than by acting on the usual causes of humiliation. So we have two variables, one psychological and one biological, which are jointly causes of later depression.

In this situation it is natural to ask: What is the mechanism by which these variables jointly cause later depression? The radical suggestion I want to consider is that there may be no mechanism. Explanation by means of mechanisms must bottom out somewhere, and then we are left with the bare facts about what would happen under interventions. At the moment, the empirical data show only that both psychological and biological variables are in general relevant to psychological outcomes. There is no empirical support for the idea that all causation that involves both psychological and biological variables bringing about a psychological outcome must be sustained by biological mechanisms. In particular, there is no reason to suppose that a comprehensive set of control variables for depression will ever be found at the biological level. It may be that the control variables for depression will always include psychological as well as biological variables.

For anyone familiar with vision science, the ubiquity of something like Marr's three levels of computation, algorithm and implementation may seem to provide a pattern that has been so successful that it's application ought to be pursued across the board. Scientists working on vision move back and forth between the cognitive level and the level of biological mechanism so seamlessly that in vision science, doing without the level of biological mechanism is almost unimaginable. But while that is certainly so for vision science, it depends on quite special features of the area that do not hold for psychological causation in general.

To explain what these special features are, I want to introduce the notion of the 'robustness' of a variable. The idea here is that if a variable does play a self-standing role in some causal process, it ought also to play a role in endlessly many other causal processes. For example, consider the so-called 'Hot Chocolate Effect': as you stir a cup of hot chocolate and the spoon sounds against the base of the cup, each successive 'ting' rises in pitch. Why is that? The usual explanation is in terms of the aeration of the liquid. As you stir, trapped air bubbles are released from the liquid and it becomes stiffer. The more rigid a substance, the faster sound travels through it. Hence, the pitch of the sound goes up (Crawford 1982). This explanation appeals to a variable, 'aeration'. Now this variable does not figure only the explanation of the hot chocolate effect. There are endlessly many ways in which you can get at the air bubbles trapped in a liquid. They are affected by the temperature of the liquid being poured into the contained, the speed at which it is poured, and they show up in as simple a way as the visible clouding of the liquid. This is what I mean by the 'robustness' of the variable: it shows up in endlessly many different causal processes and so can be investigated in endlessly many different ways.

Now consider the kinds of variables appealed to in information-processing accounts of vision. Vision is generally thought to be modular, in something like the sense of Fodor (1983; cf. Coltheart 1999). So the variables appealed to in explaining, for instance, the finer points of motion perception, or color perception, are being used to explain processing going on within a module. Now the cognitive variables – 'wavelength pattern X at place p', for instance – that are used in this kind of explanation really are internal to the characterization of the processing in a single module. What gives the brain

states the contents they have is their role in the processing within a particular modular system. It makes no sense to ask, 'What is the representational content of that cell-firing?' outside the context of inquiry into the processing going on in some particular module. For that reason, the cognitive variables appealed to in an account of some aspect of visual information-processing cannot be allowed to take on a life of their own. As purely cognitive variables, it would make no sense to suppose that the very cognitive variable that is playing a causal role in the processing going on in one module could also be playing a role in the processing going on in some other module: the determination of the content of a cognitive state here is always internal to the working of one particular module or other. The whole situation here is in sharp contrast to the appeal to aeration in explaining the hot chocolate effect, where one and the same variable can evidently be figuring in a whole sequence of quite different processes. In that sense, then, the cognitive variables appealed to in the psychology of vision are not robust. That is why we have the seamless moving back and forth between these variables and biological mechanisms. For the physiological variables are of course robust, and can be investigated through their roles in endlessly many different processes. In contrast, we give a cognitive characterization of the physiology only when we are considering the working of some one modular system.

I think that this point about robustness explains why we cannot, in vision science generally, make sense of the idea of cognitive explanation without biological mechanisms. But the point evidently does not generalize to every psychological variable. Humiliation, for example, is evidently robust. The degree to which you have been humiliated shows up in very many different causal processes. So too with the variables

of rational psychology. A particular desire may figure in causal process after causal process, leading from endlessly many different inputs to endlessly many different outputs. Your attentive awareness of an object before you may be caused by anything from it's having suddenly lit up to its having been the target of years of search, and it may play a role in processes as diverse as the starting of a train of thought and the fading of a smile. So these personal-level variables are, in general, robust. We can, therefore, appeal to them in causal explanation without having to look for the robust biological variables which might underlie them.

There may be such variables. It may be that, in the end, it will turn out that the most effective control variables for psychological outcomes in human beings are one and all biological. But at the moment, we have no evidence to support such a conclusion. At the moment, what we find are more and more biological variables working together with ever-better understood psychological variables to yield psychological outcomes. One great merit of an interventionist approach to causation in psychology, it seems to me, is that it acknowledges the possibility that this may be the right picture. We are not obliged to force the empirical findings to yield up biological mechanisms where there may be none.

ACKNOWLEDGEMENTS

This paper was begun while on leave a Stanford's Center for Advanced Study in the Behavioral Sciences, and I am deeply indebted to Alison Gopnik and Thomas Richardson for many discussions of these problems. I also had many helpful discussions with Ken Kendler. This paper was presented to a CASBS workshop on Causal Reasoning, and benefited from many helpful comments there. Thanks also to the Institute of Cognitive and Brain Sciences in Berkeley.

BIBLIOGRAPHY

Crawford, Frank S. 1982. 'The Hot Chocolate Effect'. *American Journal of Physics* 50, 398-404.

Harvey, A.G. (in press). 'Unwanted Intrusive Thought in Insomnia'. In David A. Clark (Ed.). *The Nature and Treatment of Unwanted Intrusive Thoughts in Clinical Disorders*. NY: Guilford Press.

Hausman, Daniel and James Woodward. 1999. 'Independence, Invariance, and the Causal Markov Condition'. *British Journal for the Philosophy of Science* 50, 521-83.

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'. *Archives of General Psychiatry* 60, 789-796.

Kendler, Kenneth S., Jonathan W. Kuhn, Jen Vittum, Carol A. Prescott, and Brien Riley. In press. 'The Interaction of Stressful Life Events and a Serotonin Transporter Polymorphism in the Prediction of Episodes of Major Depression: A Replication'.

Archives of General Psychiatry.

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

Pearl, Judea. 2000. *Causation*. Cambridge: Cambridge 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 Generalizations, Part 1: A Counterfactual Account'. *Nous* 37, 1-24.