Draft of August 2004
Forthcoming in *Russell's Republic*, eds. R. Corry and H. Price, OUP.

CAUSATION IN CONTEXT

Peter Menzies

Macquarie University

1. INTRODUCTION

Bertrand Russell (1917) argued that the concept of cause should be extruded from the philosophical vocabulary because it is inextricably bound up with misleading connotations.

> All philosophers, of every school, imagine that causation is one of the fundamental axioms or postulates of science, yet oddly enough, in advanced sciences such as gravitational astronomy, the word 'cause' never occurs.…the reason why physics has ceased to look for causes is that, in fact, there are no such things." (1917: p. 180)

In this paper I am more interested in the target of Russell's arguments than their cogency. Eaxctly what doctrine was he criticizing? Though he singles out several philosophers for their crude formulations of the 'law of causality', Russell does not explicitly state the target of his criticisms. Nonetheless, it is reasonable, I think, to see Russell as attacking a doctrine that might be called *causal realism*— the doctrine that causation is a feature of objective or mind-independent reality.

Actually, this is not a single doctrine, but rather a family of doctrines, varying according to what are taken to be the fundamental constituents of 'objective reality'. The particular version that Russell seems to be criticizing is one that assumes that causal relations are among the basic constituents of reality described by our most advanced physical theories. His criticism in the passage above is that this cannot be so since the advanced physical sciences do not make use of the causal concept. Put this simply, the criticism is not completely persuasive. For a defender of the causal concept could well argue that, even though causation is not explicitly mentioned in fundamental physics, it is implicitly present in the picture of reality given in fundamental physics, since causal relations supervene on the pattern of fundamental physical facts and physical laws. This more sophisticated doctrine is one Russell certainly never formulated because he did not have the concept of supervenience to hand.

In the period since Russell wrote, causal realism has become philosophical orthodoxy. The currently popular versions state that causal relations supervene on objective, mind-independent structures, though they differ with respect to what these structures are. For example, regularity theorists like J.L. Mackie (1974) think that causal relations supervene on patterns of regularities in occurrent fact. Counterfactual theorists like David Lewis (1973a, 2000) take causal relations to supervene on a network of events ordered by transitive, asymmetric relations of counterfactual dependence. Probabilistic theorists like Paul Humphreys (1989) and Ellery Eells (1991) believe that causal relations supervene on relations of probabilistic dependence between events. Process theorists Wesley Salmon (1994) and Phil Dowe (2000) believe that causal relations supervene on patterns of causal processes and interactions that involve the conservation or exchange of physical quantities. Though these theories differ in detail, they all subscribe to the doctrine that causal relations depend completely on a substructure of mind-independent relations. To be sure, a commitment to the existence of this objective

substructure often goes with a grudging admission of some minor pragmatic elements in the causal concept. For example, most causal realists are prepared to allow that pragmatic principles of 'invidious selection', as Lewis calls them, govern the way in which we select as 'the cause' a salient part of the vast network of events leading up to an event. However, setting aside such minor pragmatic complications, they claim that the causal concept has a completely objective truth-conditions which can be stated in terms of conditions holding of the mind-independent substructure.

Causal realism is the target of this paper, as it was of Russell's paper. One way to criticize this doctrine would be to show that there are features of causal claims that do not map onto any kind of objective substructure of events and relations. This kind of criticism would have to proceed on a case by case basis, as there are many versions of causal realism, differing in what they take to be the basic substructure of causation. Here I shall adopt a more general strategy. What I attempt to show is that the concept of causation is context-sensitive. Take any objective substructure of events and relations, whatever it may be, this pattern cannot determine the truth-conditions of a causal judgement, because its truth-value can vary from one context to another, depending on how a certain contextual parameter is set. The very same pattern of objective relations, viewed from within one context, may support a causal judgement, but, viewed in another context, may fail to support the judgement.

This style of argument is familiar in the case of other context-sensitive expressions. For example, Peter Unger (1975) has argued that the concept of flatness is sensitive to a certain contextual parameter that might be called a standard of flatness. This standard may vary from one context to another so the very same surface may truly be said to be flat in one context but not in another. The observation that the knowledge concept is context-sensitive in this way is at the heart of contextual theories of knowledge (de Rose 1992; Lewis 1996). I shall argue that the causal concept is subject to a similar kind of context-sensitivity. The objective facts of a situation do not determine whether one event

causes another. No more so than the objective facts about the evenness of a surface determine its flatness, or the objective facts about the kind of evidence possessed by a person determines whether that person has knowledge. In all these cases, the truth of an attribution of causation, flatness or knowledge depends on how a contextual parameter is set.

This conclusion about causation might be seen as supporting Russell's position against causal realism. If the causal concept is so riddled with context-sensitivity, then it has no use in providing an account of objective reality. (Bas van Fraassen 1980 argues for this conclusion.) Such an *error theory* about causation accepts the causal realist's characterization of reality as mind-independent, but denies the causal realist's claim that causation is part of this objective reality. But contextualism about the causal concept does not necessarily support an error theory. For another way to dispute causal realism is to reject at the outset the realist's characterization of reality in terms of certain privileged mind-independent facts. This relatively unfamiliar position might be called *perspectival realism*. The perspectival realist acknowledges that the truth-value of causal judgements does not depend entirely on the mind-independent structures. The context-sensitive character of causal judgements indicates that their truth value is perspective-relative. Nonetheless, this does not detract from the reality of the causal relations they describe. This relatively unexplored position represents, in my view, a very promising alternative to an error theory about causation. However, in this paper I will not try to adjudicate the merits of these two positions opposed to causal realism.

2. EVIDENCE FOR CONTEXTUALISM

Below I cite some evidence in support of the view that the truth-conditions of causal statements are context-sensitive. The examples I cite as evidence are, for the most part, familiar from the philosophical literature on causation. They have been discussed as counterexamples to regularity,

counterfactual or probabilistic theories of causation—the currently popular forms of causal realism. However, my interest in them lies in the fact that they demonstrate that the truth-value of causal statements can vary from one context to another. Even when the patterns of regularities, counterfactuals and probabilistic dependence are held constant, the truth-value of the causal statements can vary from one context to another.

(A) **The Indian Famine**

Pre-theoretically, we draw a distinction between causes and background conditions. The context-sensitivity of this distinction has been discussed at length by H.L.A. Hart and A. Honoré (1985).

> The cause of a great famine in India may be identified by an Indian peasant as the drought, but the World Food Authority may identify the Indian Government's failure to build up food reserves as the cause and the drought as a mere condition. (Hart and Honoré 1985: pp.35-6)

In one context, it is appropriate to judge that the following judgement is true:

> (1) The drought caused the famine and the failure to stockpile food reserves `was a mere condition of the causation.

Yet in another context a contrary statement seems to be true:

> (2) The failure to stockpile food reserves caused the famine and the drought was a mere condition of the causation.

This variation in judgement occurs despite the fact that the same regularities, counterfactuals, and probabilistic dependences hold in both contexts. For example, it is true in both contexts that if the drought had not occurred or if the government had stockpiled food, the famine would not have occurred.

(B) The Cricket Ball and Window

Michael McDermott (1995) describes the following example:

> A cricket ball is hit with substantial force towards a window. A fielder reaches out and catches the ball. The next thing along in the ball's direction of motion is a solid brick wall. Beyond that is the window. Did the fielder's catch prevent the ball hitting the window? (McDermott 1995: p.525)

Several causal judgements about this situation appear reasonable. On the one hand, because the wall would have prevented the ball from hitting the window even if the fielder had not caught the ball, we are inclined to judge that:

> (3) The fielder's catch did not prevent the ball from hitting the window.

On the other hand, because the ball was actually intercepted by the fielder and not the wall, we are inclined to judge that:

(4) The fielder's catch prevented the ball from hitting the window .

Again note that our judgements vary from one context to another even though the regularities, counterfactuals, and probabilistic dependences remain constant. For example, irrespective of whether we judge that the fielder prevented the ball from hitting the window or not, it is true that if the fielder not acted, the ball would not have hit the window, and if the wall had not been there and the fielder not acted, the ball would have hit the window.

John Collins (2000) notes that our intuitions in preemptive prevention cases of this kind can vary depending upon the nature of the backup preventer. If it is transient thing like a second fielder, we are more inclined to judge that the first fielder's catch prevented the window from shattering. If it is a permanent thing like a wall, we are less inclined to make this judgement. Finally, if the window is on the moon (so that the earth's gravitational field is a very permanent backup preventer), we are very disinclined to make this judgement. Following Collins, Lewis (2000) concludes that our judgements of causation depend upon which possibilities we deem to be too far farfetched. It is not so far-fetched that both the fielders would miss the ball, somewhat more so that the ball would avoid the wall to smash the window, and absurdly far-fetched to suppose that the ball could evade the earth's gravitational field.

(C) Contraceptive Pills and Thrombosis

Our third example was originally proposed by Hesslow (1976) as a counterexample to probabilistic theories of causation. He assumed an indeterministic setting, but I shall adapt the example to a deterministic setting.

> Betty is a young, fertile, sexually active woman who is capable of becoming pregnant. She takes contraceptive pills, which are known to cause thrombosis among women with a certain causal factor *X*. As it turns out, Betty has the factor *X* and so develops thrombosis. Let us assume that as she is taking the pill, she will avoid pregnancy, but if she were not to take the pill or use any other contraceptive method, she would become pregnant. It is a good thing that she will avoid pregnancy because when women with factor *X* become pregnant they inevitably get thrombosis.

It would seem that there are two causal judgements we can make about this case. On the one hand, since Betty did not become pregnant her getting thrombosis must be due to her taking the contraceptive pill. So it seems straightforwardly true that:

> (5) Betty's taking contraceptive pills caused her thrombosis.

On the other hand, taking contraceptive pills has a negative effect on thrombosis in women with the factor *X*, since taking contraceptive pills prevents pregnancy, which would otherwise cause them to get thrombosis. So it seems plausible to judge that:

> (6) Betty got thrombosis despite the fact that she took the contraceptive pill.

But finally because these two causal effects cancel each other out, it seems plausible to judge that:

> (7) Betty's taking contraceptive pills made no overall difference to her getting thrombosis, since she would have got thrombosis whether or not she had taken the pills.

Christopher Hitchcock (2001b) cites this example as illustrating an ambiguity in the concept of causation, which he describes in the following terms. The consumption of oral contraceptives affects a woman's developing thrombosis along at least two different routes. By analogy with the concepts of net and component forces in Newtonian mechanics, he says that contraceptive pills have two distinct effects upon thrombosis. Along one route, the one that bypasses pregnancy, the effect of contraceptive pills on thrombosis is positive, as reflected in judgement (5). Along the other route—the one that includes pregnancy or its absence—the component effect is negative: by preventing pregnancy, contraceptive pills prevent thrombosis, which is reflected in judgement (6). However, the net effect of contraceptive pills on thrombosis is neutral, as reflected in judgement (7). So he says that when we are asked whether birth control pills cause thrombosis, we can interpret this as a question about one or the other component effect, or about the net effect.

(D) The Golfer

The following example, due originally to Deborah Rosen, has been discussed at length by Wesley Salmon (1984, chapter 7). The counterexample was intended to be a counterexample to probabilistic theories of causation and is usually presented in an indeterministic setting. However, I will present the example under the assumption that causation is deterministic.

> An experienced golfer is about to drive his ball onto the green. Given his position, his level of skill and the prevailing conditions, he has an excellent chance of his holing the ball. Indeed, the only crucial variable is the angle and force of his drive: if these are within his normal range, his chance of holing out is one. But, as it happens, the player is tense and slices the ball, which veers

away from the green. But then ball the hits a tree near the green and bounces back onto the green and into the cup.

Again, we are inclined to make apparently inconsistent causal judgements about this situation. On the one hand, since we can trace a causal pathway from the golfer's slicing the ball to its falling into the cup, we are inclined to judge that:

(8) The golfer's slicing the ball caused the ball to fall into the cup.

On the other hand, since the ball was veering away from the green when it hit the tree, we are inclined to think that:

(9) The ball holed out despite the fact that the golfer sliced the ball.

One way to reconcile this apparent conflict would be to follow Elliott Sober (1984b) in saying that while statement (8) is true of token causation, statement (9) reflects a judgement that is true of type causation: a golfer's slicing the ball so that it veers away from the green generally hinders the ball from holing out. If we suppose that type and token causal claims have different semantics, we could then reconcile the apparent conflict between these causal statements. But Hitchcock (1993) has argued that this diagnosis cannot be correct, as statement (9) describes a token causal relation just as much as statement (8).

Clearly, we make these various token causal judgements because we find different features of the example salient in different contexts. Nonetheless, no matter which feature is salient in a given context,

the same pattern of regularities, counterfactuals and probabilistic dependence hold true of the situation described in the example. For example, it is true that if the golfer had not sliced the ball, the ball would have holed out; and also true that if the ball had not hit the tree, it would have veered away from the green; and so on.

3. TWO ORTHODOX RESPONSES: AMBIGUITY AND PRAGMATICS

Most philosophers find it hard to accept the conclusion that our judgements about token causation are essentially context-sensitive. In response to examples such as those of the last section, they tend to adopt one or other of two responses.

One response is to say that such examples do not impugn the idea that causal claims have context-invariant truth conditions, but rather demonstrate ambiguities in the concept of causation. This is the line of thought pursued by Hitchcock (2003), which also discusses many of the examples of the last section. According to Hitchcock, it is a mistake to look for truth-conditions for *the* concept of causation because there is simply no single unitary concept: such examples as those cited demonstrate that the verb 'cause' has many different meanings, with each meaning having its own context-invariant truth-conditions.

Contra Hitchcock, it seems to me to be remarkably implausible that the verb 'cause' should be a homonym like the verb 'bank', especially so in view of many different meanings that would have to be postulated for the verb. In the last section we considered four different, unrelated kinds of examples. If we had to posit a new ambiguity for each example, we would have four orthogonal ambiguities. However, it is implausible to suppose that our judgements about these examples depend on a sophisticated mastery of four different kinds of ambiguity. Indeed these are only a small handful of the plethora of examples that could be used to illustrate the way our judgements about causation display

uncertainty and ambivalence. These considerations, taken by themselves, are hardly conclusive, I concede. But they become more compelling when one shows, as I aim to do later in this paper, that a single set of truth-conditions, albeit ones that are relativized to a contextual parameter, can explain all the different judgements prompted by these examples.

There is another standard response to examples like those of the last section that seem to illustrate the context-sensitivity of the causal concept. It is to say that there is a univocal concept of causation that has context-invariant truth-conditions and that the different uses of the causal concept in the examples are to be explained in terms of context-sensitive pragmatic principles operating on this univocal context-invariant concept. The univocal concept is not straightforwardly evident in ordinary usage because it is partially disguised by the operation of the pragmatic principles of conversation. We can, however, recover this context-invariant concept by deliberately suspending the pragmatic principles in our philosophical discourse. Sometimes this view is expressed by using separate words for the context-invariant concept and for the compound concept that results from the operation of the pragmatic principles on the context-invariant concept. Under this regimentation, the former concept is called ‘causation proper’ and the latter is called ‘causal explanation’.

We can only assess the merits of this view by investigating whether it is successful in explaining all the diverse causal judgements we are disposed to make. To be sure, the view seems to be successful in providing a plausible explanation of the causes and conditions distinction. (See, for example, van Fraassen 1980: chapter 5.) A rough outline of the explanation goes like this. Assume that there is a network of objective causes leading up to any given event. This network consists of all those events related to the given event by some objective relation of the kind embraced by causal realists. The network will be vast, especially if the temporal ordering of events is dense, the structuring relation is transitive, and non-occurrences as well as occurrences are allowed. It is crucial that judgements about

the objective causes located in this network display no context-sensitivity. In contrast, judgements about what causally explains the event in question are highly context-dependent, as they depend on the kind of explanatory question being asked. More specifically, we might give different causal explanations of the same event in different contexts because those contexts pose different contrastive why-questions. For example, in connection with the example about the Indian famine, one might ask: Why did the famine occur *at this time* rather than some other time? A good answer to this question is to mention the drought that differentiates the present time from other times. The failure of the government to stockpile reserves of food is a mere background condition that is common to all the times and so not counted as a differentiating factor. Alternatively, one might ask: Why did the famine occur *in India*, rather than in other countries that frequently experience droughts? A good answer to this question is to mention the failure of the Indian government to build up reserves of food, which distinguishes that country from others. The occurrence of a drought is a mere condition that is common to all the different countries and so does not differentiate between them.

Perhaps this account in terms of the relativity of causal explanations to contrastive why-questions can be spelt out in detail to explain the distinction between causes and conditions. (However, I express doubts about this in Menzies 2004.) Nonetheless, the account does not seem applicable to any of the other examples cited above. These other examples seem to involve a kind of context-sensitivity that cannot be explained in terms of the relativity of causal explanations to contrastive why-questions. A systematic explanation of all these examples is needed. Until it is provided, the view that the context-sensitivity of the causal concept can be explained in terms of pragmatic principles operating on a univocal context-invariant concept is nothing more than a promissory note.

4. CAUSES AS DIFFERENCE MAKERS

Many different approaches to causation—for example, the regularity, counterfactual and probabilistic approaches —draw their inspiration from the idea that a cause is something that makes a difference to its effects, though these approaches articulate the idea in slightly different ways. What is the precise content of this idea? It is useful to try to express its content as clearly as possible to gain insight into how to articulate it most informatively. In my view, the philosophers who have best expressed the central idea that causes are difference-makers are Hart and Honoré:

> Human action in the simple cases, where we produce some desired effect by the manipulation of an object in our environment, is an interference in the natural course of events which *makes a difference* in the way these develop. In an almost literal sense, such an interference by human action is an intervention or intrusion of one thing upon a distinct kind of thing. Common experience teaches us that, left to themselves, the things we manipulate, since they have a 'nature' or characteristic way of behaving, would persist in states or exhibit changes different from those we have learnt to bring about in them by our manipulation. The notion that a cause is essentially something which interferes with or intervenes in the course of events which normally take place, is central to the commonsense concept of cause….Analogies with the interference by human beings with the natural course of events in part control, even in cases where there is literally no human intervention, what is identified as the cause of some occurrence; the cause, though not a literal intervention, is a difference to the normal course which accounts for the difference in outcome. (1985: p.29)

There seem to be three elements to Hart and Honoré's model of the way the commonsense causal concept works. The first element, which is implicit in Hart and Honoré's description, is that the

application of the causal concept to a particular situation depends upon conceptualizing the situation as involving a certain kind of system. In applying the causal concept to a particular situation, we abstract and generalize by interpreting the situation in terms of the way a particular kind of system behaves. The second element is that we typically suppose that systems of the given kind, when left to themselves, display a characteristic way of behaving. In other words, we typically suppose that systems of the given kind follow a course of evolution that is 'normal' or 'natural' for systems of that kind when they are not subject to outside interference. The third and final element is that, when the system actually exemplified in the particular situation has deviated from its normal course of evolution, we search for something that made the difference — an event or state that is analogous to an external intervention or intrusion into the system.

In my view, Hart and Honoré's model captures one important application of the causal concept, but not the only application. (For discussion of another important application see Menzies 2004, 2005.). However, to simplify my exposition, I shall focus on this application, and try to articulate it precisely so as to reveal the different ways in which it is sensitive to context. In the remainder of this section I shall try to articulate the application of the causal concept described by Hart and Honoré within the structural equations (SE) framework. Though there is a long tradition of the use of this framework in the social sciences, especially econometrics, the state-of-the-art presentation of the framework is Judea Pearl's *Causality* (2000). I shall take over the essentials of Pearl's framework, as it has been expounded by Hitchcock (2001) and James Woodward (2004: chapter 2). However, I shall introduce modifications of my own to this framework which make my treatment of token causation quite different from those of Pearl, Hitchcock and Woodward.

One distinctive feature of the SE framework is that it relativizes the truth of a token causal claim about a particular situation to a causal model. This relativization corresponds to the first element of the

Hart and Honoré's picture of the way the causal concept is applied. A causal model specifies the kind of system in terms of which we conceptualize the causal structure of the particular situation. Informally, a causal model represents a certain kind of system in terms of a set of variables representing the relevant dimensions of change for systems of the given kind and in terms of a set of generalizations governing the behaviour of systems of the given kind. More formally, a causal model is an ordered triple $<U, V, E>$. Here U is a set of exogenous variables whose values are determined by factors outside the model; V is a set of endogenous variables whose values are determined by factors within the model; and E is a set of structural equations.

The set E contains a structural equation for each variable, which appears on the left-hand side of its equation. The form of the equation depends on whether it is an exogenous or endogenous variable. If it is an exogenous variable, the equation simply states its actual value. But if it is an endogenous variable, the equation takes the form:

$$Y = f_Y(X_1, \ldots, X_n),$$

where $X_1, \ldots, X_1$ are all and only the variables from the sets U and V that play a role in determining the value of Y. It is important to note that the structural equations are not standard symmetric equations. In these equations side matters: the values of the variable on the left-hand side are *determined* by the values of the variables on the right-hand side. Different theorists understand the structural equations in slightly different ways. Pearl understands them to represent the basic causal mechanisms governing the behaviour of the system of the given kind, with each equation representing a distinct causal mechanism.

A causal model and its set of structural equations can be depicted in a graphical representation. The variables in the sets *U* and *V* form the nodes of a graph. These nodes are connected by directed edges according to the following rule: an edge is drawn from *X* to *Y* if and only if *X* appears in the right-hand side of the structural equation for *Y*; in other words, if and only if the values of *X* play a role in determining the values of *Y*. In this case, *X* is said to be *a parent of Y*. An exogenous variable is one without any parent in the graph for its model.

In relativizing causal claims to a causal model, the SE framework clearly introduces several degrees of freedom in representing the causal structure of a particular situation. As we shall see, this will be crucial to explaining the context-sensitivity of causal claims. However, it is important to emphasize that this relativization should not be seen as introducing an excessive degree of subjectivity. For example, philosophers who study causal models sometimes remark that the causal structure implied by a model depends on the choices made by the modeler — for example, choices about how to represent the situation in terms of variables, and about whether to represent them as binary or many-valued variables. In contrast, Pearl strongly dissents from such remarks. As I read him, he believes that it is a completely objective matter how a certain kind of system is to be represented in a model. The set of variables *U* and *V* of a model express all and only the objective dimensions of change for systems of the given kind, where these are joints in nature, discovered rather than constructed by us. He has correspondingly objectivist views about the set of structural equations *E* of a model. As remarked above, for him these represent the basic causal mechanisms that govern the behaviour of systems of the given kind. Their existence and nature are completely mind-independent matters that are settled on the basis of objective experimental and observational methods. I shall follow Pearl in his objectivist construal of the way models represent the kinds of systems that are implicitly invoked in causal talk.

It is probably best to explain the SE framework by means of a simple example. Let us adapt an example introduced by Hitchcock (1996) for a different purpose. (We shall consider Hitchcock's original purpose eventually.) Let us suppose that a person is given a certain drug, 'curit', in order to cure him of a disease from which he is suffering. He can be given different doses of the drug: no dose, a moderate 100mg dose, or a strong 200mg dose. The drug is known to be effective in large doses, but the cost and the risk of side-effects make it impractical to give a large dose to this patient; and so he is given a moderate dose of 100mg. As it happens, the patient recovers; and we ask 'Did taking the moderate dose make a difference to the patient's recovery?'

To answer this question within the SE framework, we have to choose a causal model that specifies the kind of system we are investigating. Let us model the patient as a human physiological system, but let us dramatically oversimplify this kind of system by supposing it can be characterized in terms of two variables: an exogenous variable C, which can take three values 0, 100mg, and 200mg corresponding to the different dosage levels; and an endogenous variable R, which takes the value 1 if the patient recovers and 0 if he does not recover. Finally, let us suppose that the structural equations of this model are as follows:

$C = 100\text{mg}$;

$R = f(C)$, where $f(C) = 1$ if $C \geq 100\text{mg}$ and 0 otherwise.

The first equation is the structural equation for the exogenous variable C. The second is the structural equation for the endogenous variable R: it states that a patient recovers if and only if he is given a dose of the drug curit greater than or equal to 100mg. It is simple matter to calculate that the actual value of the variable R is 1. The graph for this model is depicted in Figure 1:

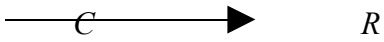

Figure 1

In order to answer the question 'Did taking the moderate dose of curit make the difference to recovery?' within the SE framework, one has to assess whether there is a counterfactual dependence between his taking the moderate dose and his recovery. For the framework construes the difference-making relation in terms of counterfactual dependence. Indeed, it is one of the nice features of this framework that it provides an elegant method for evaluating counterfactuals. To evaluate a counterfactual whose antecedent specifies the value of a variable, whether exogenous and endogenous, we simply replace the equation for the relevant variable with one that stipulates the new value of the variable, while keeping all the other equations unchanged. For example, to calculate what would have happened if the patient had been given no dose or a strong dose of curit, we simply replace the structural equation C = 100mg with either C = 0mg or C =200mg and recalculate the value of the endogenous variable R. When we do this we can see that the following counterfactuals are true:

C = 0mg □→ R = 0;

C = 200mg □→R = 1.

Within the SE framework, the action of replacing an equation by another stipulating a new value represents the way in which the value of the relevant variable might be set by an intervention from outside the system. Of course, this is a figurative way of expressing the matter. But it is hard to avoid

the use of metaphorical talk. For example, Lewis's possible worlds semantics for (non-backtracking) counterfactuals invokes a similar analogy when it stipulates that the closest deterministic worlds in which a counterfactual antecedent is true are ones in which the antecedent is realized by a miracle. The action of setting the value of a variable by a surgical intervention has special significance in the case of an endogenous variable. Replacing the equation for an endogenous variable with an equation stipulating a new value is in effect treating the endogenous variable as an exogenous one. Graphically, the edge that is directed into this variable is removed while all the other edges remain intact. As the proponents of the SE framework acknowledge, his technique for evaluating counterfactuals depends on certain non-trivial theoretical assumptions. First, it depends on the assumption that the system of equations is *modular* in the sense that one can surgically change the equation for one variable without thereby disturbing the equations for the other variables. Secondly, it depends on the assumption that the equations of a system are individually *invariant* in the sense that they continue to hold under interventions that set the values of some of their variables.

This technique for evaluating counterfactuals enables one to capture the idea of counterfactual dependence. The most natural definition within the SE framework is this:

> Definition 1: A variable Y *counterfactually depends* on a variable X in a causal model if and only if it is actually the case that $X = x$ and $Y = y$ and there exist $x' \neq x$ and $y' \neq y$ such that the result of replacing the equation for X with $X = x'$ yields $Y = y'$.

As noted above, this notion of counterfactual dependence is supposed to capture the idea of difference-making. The value of one variable *makes a difference* to the value of another variable precisely in the

sense that if the first variable had been wiggled to change its value, then the second variable would have changed its value too.

However, we now come to the problem that Hitchcock originally used the example to illustrate. Unfortunately, one cannot apply the definition of counterfactual dependence to the example at hand to determine a unique, unequivocal answer to the question whether giving the moderate dose makes a difference to the patient's recovery. For there are two different counterfactual cases that contrast with the actual case in which the patient is given the moderate 100mg: the case in which he is given no dose of curit and the case in which he is given the strong 200mg dose. The problem lies in the fact that these different contrast cases give conflicting answers to the question about counterfactual dependence. If we suppose that the hypothetical case that contrasts with the actual case is one in which the patient is given no dose of curit, then patient's recovery counterfactually depends on his taking moderate dose. However, if we suppose that the contrasting hypothetical case is one in which the patient is given the strong dose, the patient's recovery does not counterfactually depend on his taking the moderate dose. Clearly these results cannot both be correct.

The moral that Hitchcock draws from this kind of example is that we are mistaken in thinking that causation is a binary relation between a cause and effect. It is this assumption that forces us into choosing one or other result. If we think of causation as a ternary relation between a cause C and effect E relative to an alternative cause C', we can accept both results. Indeed, causal language has devices that make it possible for us to express the matter clearly. We can say the following: administering the moderate dose of curit, as opposed to no dose, made a difference to his recovery, whereas administering the moderate dose, as opposed to the strong dose, did not. Hitchcock argues that such contrastive clauses enable us to capture the extra relatum that is usually unexpressed in causal statements. Contrastive stress also serves the same purpose. For example, we might express the above

idea in this way: the patient recovered because he was given *some dose* of the drug, not because because he was given the *moderate* dose. This use of contrastive stress indicates that the alternative cause in the first case is 'no dose of the drug' whereas the alternative cause in the second case is 'the strong dose of the drug'.

I am inclined to think, however, that a more far-reaching moral can be drawn from this example. Let us reconsider the model of difference-making described by Hart and Honoré. The SE framework captures some of its main ideas. By relativizing causal discourse to a model, it captures the idea that causal claims involve an implicit relativity to a kind of system. It also captures the idea of a cause as a difference-maker in terms of notion of counterfactual dependence. On the other hand, the SE framework, in its present form, does not capture the guiding idea that a cause of some effect in a system is something *analogous to an intervention* from outside the system that makes a difference to the *normal course of evolution* of that system. I propose to introduce some modifications into the SE framework to capture these missing elements.

I shall modify the framework by allowing that the description of the model used to interpret a particular situation may fix the values of exogenous variables at non-actual values. In standard presentations of the framework, the values of the exogenous variables that are taken as the baselines for the calculation of counterfactual dependences are always the actual values of the variables. I propose a reversal of this usual order. Let us suppose that in framing the model relevant to a particular situation, we are permitted to set the values of exogenous variables at possible values to represent what I shall call default states of the system. Informally, the default values of the variables represent the *normal* or *natural* state of the system in question. I shall explain this notion in more detail when I apply it to particular examples. I shall call a causal model with its exogenous variables set at their default values *a default causal model*.

We cannot apply the usual definition of counterfactual dependence to a default causal model because the definition anchors a counterfactual dependence to the actual values of the variables. We have to redefine central notion of a difference maker as follows:

> Definition 2. A value of a variable X *makes a difference* to the value of another variable Y in a default causal model if and only if plugging in the default values of the variables in the structural equations yields $X = x$ and $Y = y$ and there exist actual values $x' \neq x$ and $y' \neq y$ such the result of replacing the equation for X with $X = x'$ yields $Y = y'$.

It is easy to see the implications of this modification of the SE framework by applying it to the curit example. Suppose we say that the normal state of the patient is one in which he receives no dose of curit. So we fix on a default model in which the variable C is given the default value 0, and we reason accordingly that the patient's taking the moderate dose of the drug made a difference to his recovery. On the other hand, if we suppose that the normal state of the patient is one where he is given the strong dose of the drug, so that the variable C is given the default value 200mg, we can reason that the patient's taking the moderate dose made no difference. In this way, relativizing the assessment of a causal claim to a default causal model is roughly similar to Hitchcock's strategy of a rephrasing causal claim as a ternary relation between a cause, an effect, and a contextually determined alternative cause.

However, it is important to note the differences between the proposed strategy and Hitchcock's strategy. One difference is that Hitchcock's strategy is presented as a stand-alone strategy intended to deal with a specific difficulty. In contrast, the strategy proposed above is meant to be part of a larger strategy for formalizing the Hart and Honoré model of causation within the SE framework. As such, the strategy gains its intelligibility and its plausibility from this larger attempt to capture the insights of

the interventionist or agency model that Hart and Honoré describe. For example, it is central to this model that a cause is seen as analogous to an intervention into a system that makes a difference to the system's normal course of evolution. Hitchcock's strategy makes no reference to this requirement on a cause. Another difference between the strategies concerns the scope of the context-sensitivity of causal claims. On Hitchcock's strategy, context points to an alternative possible cause to be contrasted with the actual cause, while, on the strategy proposed above, context points to a whole set of alternative possible worlds that realize the alternative cause. In other words, the proposed strategy involves the idea that context picks out a whole set of alternative background conditions as well as a contrasting cause. It is this insight—that context affects not just the choice of a contrast case, but also the set of alternative possible worlds that realize the contrast case—that provides the key to understanding the examples in section 2.

5. THE EXAMPLES REVISITED

Let us return to the examples described in section 2 to see how the modified SE framework can explain our various causal judgements.

(A) The Indian Famine

We might represent this example in terms of a model employing the following variables:

D = 1 if a drought occurs in India, 0 if not.

R = 1 if the government stockpiles reserves of food, 0 if not.

F = 1 if a famine occurs in India, 0 if not.

The structural equations of this model will consist of the following equations:

$D = 1$

$R = 0$

$F = D \,\&\, \sim R$

(Here logical symbols are used to represent the obvious mathematical functions on binary variables: $\sim X = 1\text{-}X$, $X \text{ v } Y = \max \{X, Y\}$; $X \,\&\, Y = \min \{X, Y\}$.) The first two equations state the values of the exogenous variables *D* and *R*. The third equation states the value of the endogenous variable *F* as a function of the variables *D* and *R*. The graph corresponding to this model is depicted in Figure 2.

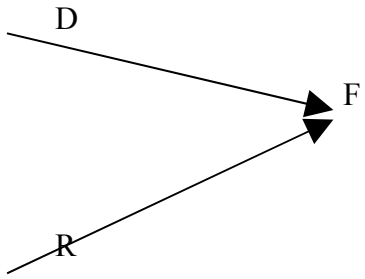

Figure 2

If we were to try to determine the difference-making relations according to the standard version of the SE framework, we would fix the exogenous variables *D* and *R* at the actual values 1 and 0; and then we would use those values as the baseline for assessing whether various counterfactual dependences hold. On this way of proceeding, it turns out that the famine counterfactually depends on the drought, but also depends on failure of the Indian government to stockpile food reserves. These counterfactual

dependences do not vary from one context to another and so are unsuitable for explaining how the drought makes the difference to the famine in one context, and the failure to stockpile food reserves makes the difference in a different context.

However, I have argued that it is necessary to introduce a modification to the SE framework to capture for an extra dimension of context-sensitivity. This modification allows us to set the values of some variables at non-actual default values and to use them as baselines for the calculation of the difference-making relations. The default values of variables represent the states of the system that are deemed to be normal or natural in some sense.

Returning to the example at hand, let us suppose that the normal situation in India is one in which there is no drought and that the Government does not stockpile food. We can create a new default model M_1 from the old one by resetting the values of the exogenous variables at $D = 0$ and $R = 0$. Applying Definition 2 to this model, we see that the drought makes a difference to the famine, but the failure to stockpile food reserves does not. As I shall show later, when the drought makes a difference to the famine in a default model M_1, the pair of causal conditionals in (10) will hold true. I index these conditionals by M_1 to signify that they are relative to a particular default causal model M_1.

(10) $(D{=}1 \mathrel{\Box\!\!\rightarrow}_{M1} F{=}1)$ & $(D{=}0 \mathrel{\Box\!\!\rightarrow}_{M1} F{=}0)$

This is just the result we would expect: the drought makes a difference with respect to the famine and the government failure to stockpile food reserves is a mere background condition that does not make a difference.

Now let us make a different assumption about the default values of the exogenous variables *D* and *R*. Let us suppose the normal situation is one in which drought regularly occurs and the government takes it upon itself to build up food reserves against the possibility of drought. So let us

assume the default values of the exogenous variables are $D = 1$ and $R = 1$ in a default model M_2. Then it is possible to see that in this new default model, the failure of the government makes a difference to the famine, but the drought does not. Consequently, the pair of causal conditionals in (11) hold true:

(11) $(D=1 \; \square\!\!\rightarrow_{M2} F=0) \; \& \; (D=0 \; \square\!\!\rightarrow_{M2} F=0)$

Once again we have the result that agrees with intuition: in this context the government failure makes a difference with respect to the famine, while the drought is a mere background condition.

(B) The Cricket Ball, Fielder and Window

We can model this example using the following variables:

$B = 1$ if a ball is flying in direction of window, 0 if not.

$F = 1$ if fielder catches the flying cricket ball, 0 if not

$W = 1$ if the wall is present, 0 if not.

$S = 1$ if cricket ball shatters the window, 0 if not.

The structural equations of the model with actual settings of exogenous variables are:

$B = 1$

$F = 1$

$W = 1$

$S = B \ \& \sim F \ \& \sim W$

The graph for this model is depicted in Figure 3.

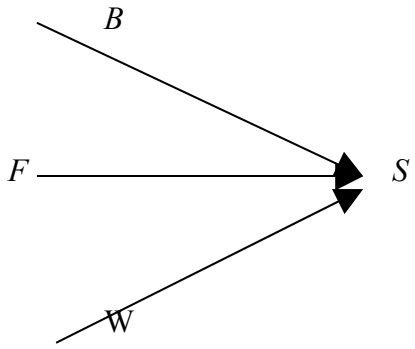

Figure 3

Let us consider how to convert this model into an appropriate default model. Let us assume that the system we are modeling is a flying-ball-plus-window-protected-by-a-wall and that its normal state is one in which a ball is flying through the air, the fielder does not catch the ball but the wall is present to protect the window. Accordingly, we set the default values of the exogenous variables at $B = 1$, $F = 0$ and $W = 1$ in a default model M_1. In this case, it is easy to check that the fielder's catching the ball makes no difference to the window's not breaking:

(12) $F=1 \ \Box\!\!\rightarrow_{M1} S = 0 \ \& \ F = 0 \ \Box\!\!\rightarrow_{M1} S=0$

In contrast, let us suppose that the system we are modeling is a flying-ball-plus-unprotected-window and that it is normal state is one in which a ball is flying through the air, the fielder does not catch the ball, and the wall is not present. (It is easier to think of such a default model the more readily we can detach the back-up preventer from the system. This is difficult to do if the back-up preventer is

a wall; and much easier to do if it is a second fielder. Compare this diagnosis with that of Collins and Lewis.) Accordingly, let us set the default values of the exogenous variables at $B = 1$, $F = 0$ and $W = 0$ in a default model M_2. Then it is easy to see that the fielder's catching the ball makes a difference to the window's not shattering, as reflected in the truth of the pair of causal conditionals:

(13) $F{=}1 \; \Box\!\!\rightarrow_{M2} S{=}0$ & $F{=}0 \; \Box\!\!\rightarrow_{M2} S{=}1$

So our intuitions about the causal structure of this situation depend on what we take to be the system in question and what we take to be the normal state of this system. We are more inclined to see the fielder's catch as preventing the ball from shattering the window when we think of the system in question as a flying-ball-plus-unprotected-window whose normal state the does not involve the presence of the wall. (See a similar diagnosis of this example in Maudlin 2004.)

(C) Contraceptive Pills and Thrombosis

It is natural to model this example using the following variables:

$X = 1$ if Betty has factor X, 0 if not.

$C = 0$ if the Betty does not take contraceptive pills, 1 if she does, and 2 if she uses other contraceptive means.

$P = 1$ if Betty is pregnant, 0 if not.

$T = 1$ if Betty gets thrombosis.

The causal model with the actual values of the exogenous variables has the following structural equations:

$X = 1$.

$C = 1$

$P = 1$ if $C = 0$ and 0 otherwise

$T = 1$ if (P=1 and X=1) or (X =1 and C =1); 0 otherwise.

The graph for this example is depicted in Figure 4.

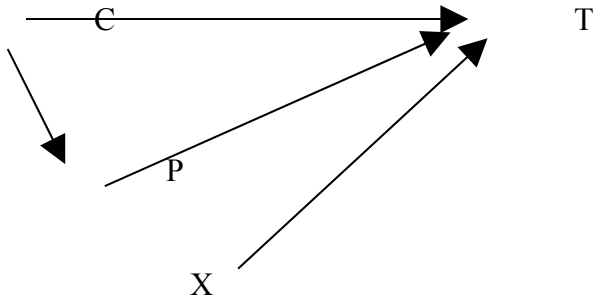

Figure 4

Now let us consider how this model can be modified in the light of different assumptions about the system being modeled and about its normal state. Suppose that the system we are modeling is a fertile-sexually-active-woman-with-factor-*X*-who-does-not-take-contraceptive-pill so that the default values of the exogenous variables of this system can be set at $X = 1$ and $C = 0$. Then it is easy to see that within the default model M_1 with these default settings, Betty's taking the pill makes no difference to her getting thrombosis. For if she were to take the pill, she would get thrombosis because she has factor *X*; and if she were not to take the pill, she would get thrombosis because she has factor *X* and would become pregnant. The following causal conditionals hold true:

(14) (C=1 $\square\!\!\rightarrow_{M1}$ T = 1) & (C = 0 $\square\!\!\rightarrow_{M1}$ T=1)

Now let us make a different assumption about the system being modeled. Let us suppose that it is a fertile-sexually-active-woman-with-factor-X-who-does-not-take-contraceptive-pill-but-uses-other-contraceptive-means so that the default values of the variables are X = 1 and C = 2. It is easy to check that in the default model M_2 with these settings, Betty's taking the contraceptive pill makes a difference to her getting thrombosis:

(15) (C=1 $\square\!\!\rightarrow_{M2}$ T = 1) & (C = 2 $\square\!\!\rightarrow_{M2}$ T=0)

This corresponds to the positive causal judgement that Betty's taking the pill caused her thrombosis.

(D) The Golfer

It is natural to model this example using the following variables:

D = 0 if the golfer does not drive at all, 1 if he drives with normal angle and force, 2 if he slices the ball.

T = 1 if tree is present near green, 0 otherwise.

R = 1 if golf ball ricochets off tree towards green, 0 otherwise

H = 1 if ball holes out, 0 otherwise

The structural equations for this model with actual values for the exogenous variables are:

$D = 2$

$T = 1$

$R = 1$ only if $D = 2$ and $T = 1$; and 0 otherwise.

$H = 1$ only if $D = 1$ or (D=2 and R=1); and 0 otherwise.

The graph for this model is depicted in Figure 5.

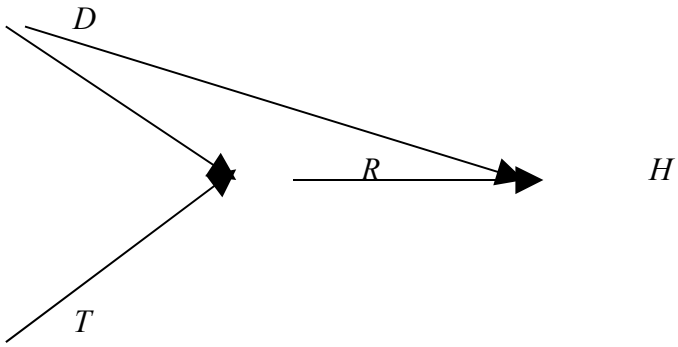

Figure 5

In order to capture the various intuitions about this example we have to convert this model into various default models. Let us model the situation as a system consisting of a golfer-plus-tree-near-green and let us suppose the normal state of this system is such that $T = 1$ and $D = 0$. Then we can deduce that golfer's slicing the ball made a difference to the ball's holing out in this default model M_1:

(16) ($D = 2 \; \Box\!\!\rightarrow_{M1} H = 1$) & ($D = 0 \; \Box\!\!\rightarrow_{M1} H = 0$)

In order to capture the intuition that in some sense the ball holed out despite the fact that the golfer sliced it, we must model the system in a different way. Suppose we think of the system as golfer-plus-green-without-tree, reflecting the idea that the tree is an accidental and not an essential part of the

system. Accordingly, let us set the default value of T to 0, keeping the default value of D at 1 in a default model M_2. Then we can see that the pair of causal conditionals hold:

(17) $(D = 2 \mathrel{\Box\!\!\rightarrow}_{M2} H = 0)$ & $(D = 1 \mathrel{\Box\!\!\rightarrow}_{M2} H = 1)$

Of course, such a despite-causal claim gets its point precisely when $D = 2$ makes a difference to $H = 0$ but in fact, due to other causal factors, it happens to be the case that $H = 1$.

6. CAUSAL CONDITIONALS

I have employed causal conditionals in the previous section to illustrate various judgements about the difference-making relation. How are these conditionals to be understood? Do they have a coherent semantics? And how do they relate to standard counterfactuals?

Let us tackle the last question first. Traditionally, philosophers have developed semantics for counterfactuals in terms of similarity relations between possible worlds. One classic treatment is David Lewis's (1973b) possible worlds semantics. (Pearl in his (2000, section 7.4) shows that the axioms of Lewis's theory follow from the axioms of his own structural semantics.) A central feature of Lewis's semantics is that it uses a system of nested spheres of possible worlds centered on the actual world. A sphere represents a set of possible worlds that are equally similar to the actual world: the smaller the sphere the more similar to the actual world are the possible worlds within it.

Built into this semantics is the Centering Principle to the effect that there is no world as similar to the actual world as the actual world itself. In terms of this system of spheres the truth condition for a counterfactual is stated as follows: $P \mathrel{\Box\!\!\rightarrow} Q$ is true if and only if Q is true in every P-world in the

smallest P-permitting sphere. It follows from the Centering Principle that $P \mathrel{\square\!\!\rightarrow} Q$ is true if P and Q are true.

I propose a modified semantics for the causal conditionals that are relevant to the assessment of the difference-making relation in the modified SE framework. This semantics differs from Lewis's in two ways. The first difference is that the similarity relation is specified by reference to a contextually salient default causal model. Such a causal model determines the relevant respects of similarity to be considered in evaluating a given counterfactual. The second difference is that the system of spheres of possible worlds need not be centered on the actual world, but on a set of what I call default worlds. I characterize default worlds as follows:

> Definition 3*:* A *default causal model* <*U*, *V*, *E*> of an actual system generates a sphere of default worlds that consists of all and only worlds *w* such that:
>
> (i) *w* contains a counterpart system of the same kind whose exogenous variables in *U* are set at their default values;
>
> (ii) *w* evolves in accordance with the structural equations in *E* without any further intervention from outside the system.

The intuitive idea is that the default worlds generated by a causal model exemplify a course of evolution that is normal or natural for a system of the given kind. More particularly, they represent the way that a system of the given kind would evolve from its initial default state without intervention or interference from outside the system. A crucial notion here, of course, is that of the default settings of the exogenous variables of a model, about which I shall say more in the next section.

Let us consider how this definition would apply to an example. What would a default world generated by a causal model for example 1, say, look like? As we have seen, the exogenous variables are *D* and *R* and their default values are determined in a context-sensitive way. We have seen that in one context in which it is normal for there to be no drought and no government stockpiling of food reserves, they have the default values of 0 and 0 respectively. The default worlds generated by this default model will be ones that evolve from these initial values in conformity with the relevant structural equation in such a way that there is no famine. However, in another context in which droughts and government stockpiling food reserves are the norm, they have the default values of 1 and 1 respectively. The default worlds generated this default model will be ones that evolve from these initial states in such a way that there is no famine.

The sphere of default worlds generated by a model is tied, in some sense, to the actual world. For worlds earn their membership in the sphere by virtue of their resemblance to the way the actual system under consideration would evolve in conformity with the structural equations. Nonetheless, it is important to note that the actual world need not itself belong to the sphere of default worlds. For these worlds represent how a normal system would evolve in conformity with the structural equations in the absence of outside intervention. In many cases, these worlds are ideal ones. The actual world may be very far from ideal in that the actual system may not be normal and its course of evolution may be affected by external interferences. For example, in the Indian famine example, the actual world is one where there is drought and no government stockpiling of food reserves, whereas the default worlds generated by the two models we considered lack one or other of these features. The sphere of default worlds within this framework count as the closest worlds to the actual world. The fact that the actual world need not belong to this sphere means that the Centering Principle fails in this framework. This

has some surprising implications for the logic of causal conditionals, including the failure of Modus Ponens, but these must be explored elsewhere.

So far we have attended to the question of which worlds count as the default worlds generated by a default causal model. But we need to provide truth-conditions for all causal conditionals, and so we need to specify which will be the closest-antecedent worlds for any causal conditional. In some cases, the antecedent of the conditional will overlap with the sphere of default worlds, and so the closest antecedent-worlds are specified as those antecedent-worlds that belong to this overlap. In other cases, however, the antecedent of the causal conditional will not overlap with the sphere of default worlds and the closest antecedent-worlds must be specified in some non-obvious way. Here I propose one way in which a default causal model might create an ordering of spheres of possible worlds, adapting an idea of Pearl's (2000: p.241)

> Definition 4: $\{S_0, ..., S_n\}$ is *a system of spheres* ordered by the model $<U, V, E>$ if and only if S_0 is the sphere of default worlds generated by the model and S_i is a sphere of worlds such that a default world in S_0 is transformed into a world in S_i by a maximum number of $\underline{i}$ interventions in the structural equations of the model.

It is easy to see that this method of ordering the spheres of possible worlds ensures that they are centered on the sphere of default worlds and are nested within each other.

On the basis of this ordering of spheres, the truth conditions for the causally relevant counterfactuals can be formulated in general terms as follows:

Definition 5: $P \mathrel{\Box\!\!\rightarrow}_M Q$ is true in the actual world relative to a system of spheres ordered by a default causal model M if and only if *Q* is true in all the *P*-worlds in the smallest *P*-permitting sphere of the system of spheres of default worlds generated by the model.

It can be shown that these definitions provide a coherent semantics for the causal conditionals used to illustrate the difference-making relations discussed in the last section.

7. DEFAULT WORLDS

The preceding discussion has focused on one method (but not the only method) for evaluating causal relations. According to this method for determining causes, we try to conceptualize a particular situation in terms of certain kind of system that has a characteristic normal course of evolution. If we are successful in doing this, and we also encounter some behaviour of the system that deviates from this normal course of evolution, we judge something to be a cause of this deviation if it has made a difference to the normal course of events with respect to the behaviour.

Many philosophers will recognize this form of reasoning as Aristotle's natural state model of scientific reasoning. Elliott Sober (1980, 1984) has written that the model provided Aristotle with a technique for explaining the great diversity found in natural objects. Within the domain of physics, there are heavy and light objects, ones that move violently and ones that do not move at all. How is one to find some order that unites and underlies all this variety? Aristotle's hypothesis was that there is a distinction between the natural state of a kind of object and the states that are not natural. The latter are produced by subjecting the object to an interfering force. In the sublunar sphere, for a heavy object to be in its natural state is for it to be located where the center of the Earth is now. But, of course, many

heavy objects fail to be there. The cause for this divergence from what is natural is that interfering forces act on the objects to prevent them from achieving their natural state. Aristotle's metaphysics reflects the kind of default reasoning that I believe is central to causal reasoning.

But this form of reasoning is not just a fossilized piece of folk science. It is common to many forms of advanced scientific reasoning. Tim Maudlin (2004) has argued that causal reasoning is especially straightforward and clear when we have laws of a particular form, illustrated by Newton's laws of motion. The first law, the law of inertia, states that a body at rest will remain at rest and a body in motion will continue in motion at a uniform speed in a straight line, unless some force is put on it. The first law specifies inertial motion, that is, how the motion of an object will evolve if nothing acts on it. The second law then specifies how the state of motion of an object will change if a force is exerted on it: it will change in the direction of the force, and proportionally to the force, and inversely proportionally to the mass of the object. The structure of these laws is especially well suited to identifying causes, Maudlin claims. Where we encounter an object deviating from its inertial state of motion (eg the Earth orbits the Sun rather than traveling at constant speed in a straight line), we look for a force that explains the deviation (eg the gravitational force produced on the Earth by the Sun). In this form of reasoning the second law is parasitic on the first: the first specifies the inertial motion deviation from which requires explanation in terms of a force. If the inertial motion of the Earth were to orbit the Sun, its actual motion would not require a cause in the form of a force.

Maudlin argues that this form of causal reasoning is more general and not peculiar to physics. He writes:

> In judging causes, we try to carve up the situation into systems that can be assigned inertial behavior (behavior that can be expected if nothing interferes) along with at least a partial

> specification of the sorts of things that can disturb the inertial behavior, analogous to the Newtonian forces that disturb inertial motion. Let us call the things that can disturb inertial behavior 'threats': they are objects or events that have the power —if they interact in the right way—to deflect the system from its inertial trajectory. We then think about how the situation will evolve by expecting inertial behavior unless there is interaction with a threat, in which case we see how the threat will change behavior. (2004: p.436)

Evidently, Maudlin is making the same point I made at the beginning of this section. Where he talks of the inertial behaviour of a system, I talk of its normal course of evolution or the default worlds that exemplify its normal behaviour; and where he discusses threats that explain changes from the inertial behaviour, I discuss interventions or intrusion in the system that make a difference to its behaviour.

It would be good to have a detailed account of the notion of default world generated by a model of system and the correlative notion of the default setting of the exogenous variables of a causal model, especially in view of the reliance of our discussion on these notions. In very general terms, the default values of the exogenous variables represent the initial state of a system of a given kind that is normal or to be expected or is taken for granted because it requires no explanation; and the corresponding default worlds represent the normal course of evolution of the system from this initial state— normal in the sense of conforming to the structural equations in a way that is free from intervention from outside the system.

However, it is very hard to frame more detailed, positive characterizations of these notions since they are so subject to subtle idiosyncratic contextual cues, as we have seen in the examples above. Nonetheless, it is possible to make some very rough and ready generalizations about how the default worlds are determined for a system.

First, known laws or regularities clearly influence the expectations of what is the normal course of evolution for a system of the given kind. (See Toulmin 1961: chapters 3 and 4). This applies in both everyday and scientific explanations. If every car passing down the street in front of my house has been reasonably quiet, then the expectation based on that regularity will determine what I think calls for explanation. A car backfiring, for example, will be something that is anomalous with respect to that expectation and will require explanation. To take a scientific example: if every planet in the solar system has been observed to conform to an elliptical orbit predicted by Newton's laws, then that expectation will form the basis of what counts as an anomalous phenomenon. When a planet is observed to deviate from its predicted elliptical orbit, an explanation will be sought in terms of the gravitational influence of a yet-to-be discovered planet.

Secondly, the default worlds are not restricted to the regular course of events unaffected by human intervention. Because nature can be harmful unless we intervene, we have developed customary techniques, procedures and routines to counteract such harm. (See Hart and Honore 1985.) These become a second nature. For example, the effect of drought is regularly neutralized by government precautions in conserving water; disease is neutralized by inoculation; rain by the use of umbrellas. When such procedures are established, they can determine the default worlds for the relevant system. When some harm occurs in violation of the expectations set up, a cause is often identified as an omission or failure on the part of some agent to carry out the neutralizing procedures, as the example of the famine illustrates.

Finally, another factor that can determine the default worlds for a given kind of system is what is regarded as the proper functioning of that system. For example, doctors have expectations about the normal or healthy functioning of human physiology based on their beliefs about the proper functions of the body. Their search for the causes of deviations from healthy functioning will also be influenced by

these expectations as well. Similarly, biologists' expectations of the default or normal course of development of a biological trait will be influenced by their views of the adaptive function of the trait.

I am sure that it is possible to add to and improve upon these rough and ready generalizations. There may, indeed, be laws about the way in which the human mind forms conceptions of the inertial behaviour of systems and the default worlds they exemplify. However, I doubt whether any of such laws, which might improve on the rough and ready generalizations given above, could provide a general *positive* characterization of inertial behaviour that applies to every kind of system. For I suspect that the notion of the inertial behaviour or a default world for a kind of system has to be understood *negatively*: the inertial behaviour of a Newtonian system is simply the behaviour of the system when *no external force interferes*; and more generally, the default world for a system exemplifies the behaviour it would display if *there were no external causes at work*. It might be thought to be a deeply unsatisfying feature of this model of causal reasoning that no positive characterization can be provided of the central notion of inertial behaviour or default world. Nonetheless, it would be inadvisable to reject this model for this reason. For as we have seen, the kind of default reasoning I believe to central to causal reasoning is also common to many of our best scientific theories. As already pointed out, this form of causal reasoning is embedded in Newtonian mechanics in the form of its first and second laws. Sober (1980; 1984) also points out that the same model of reasoning is used in population genetics under the rubric of the Hardy-Weinberg law. This law specifies the equilibrium state for the frequencies of genotypes in a population when the evolutionary forces of mutation, migration, selection and drift are not at work. Even in general relativity, the geometry of space-time specifies a set of geodesics along which an object will move as long as it is not subjected to a force. Consequently, the rejection of the kind of default causal reasoning described in this paper would necessitate the rejection of the reasoning embedded in some of our best scientific theories.

REFERENCES

Collins, J. 2000. "Premptive Prevention", *Journal of Philosophy*, 97:223-234.

De Rose, K. 1992. "Contextualism and Knowledge Attributions", *Philosophy and Phenomenological Research*, 52, pp.913-929.

Dowe. P. 2000. *Physical Causation*. Cambridge: Cambridge University Press.

Eells, E. 1991. *Probabilistic Causation*. Cambridge: Cambridge University Press.

Hart, H. and Honoré, T. 1985. *Causation in the Law*, 2ed. Oxford: Oxford University Press.

Hesslow, G. 1976. "Discussion: Two Notes on the Probabilistic Approach to Causality", *Philosophy of Science*, 43: 290-292.

Hitchcock, C. 1993. "A Generalized Probabilistic Theory of Causal Relevance", *Synthese*, 97:335-64.

Hitchcock, 1996. "Farewell to Binary Causation", *Canadian Journal of Philosophy*, 26:267-282.

Hitchcock, C. 2001a. "The Intransitivity of Causation Revealed in Equations and Graphs", *Journal of Philosophy*, 98: 335-64.

Hitchcock, C. 2001b. "A Tale of Two Effects". *The Philosophical Review*, :361-396.

Hitchcock, C. 2003. "Of Humean Bondage", *The British Journal for the Philosophy of Science*,

Humphreys, P. 1989. *The Chances of Explanation*. Princeton: Princeton University Press.

Lewis, D. 1973a. 'Causation", *Journal of Philosophy*, 70: pp.556-567.

Lewis, D. 1973b. *Counterfactuals*. Oxford: Blackwells Publishing.

Lewis, D. 2000. "Causation as Influence", *Journal of Philosophy*, 97: pp182-97.

Mackie, J. 1974. *The Cement of the Universe*, Oxford: Oxford University Press.

Maudlin, T. 2004. "Causation, Counterfactuals, and the Third factor", in J. Collins, N. Hall. and L. Paul (eds.), *Causation and Counterfactuals*. Cambridge, Mass.: MIT Press: pp.419-443.

McDermott, M. 1995. "Redundant Causation", *British Journal for the Philosophy of Science*, 40: 523-544.

Menzies, P. 2004. "Difference Making in Context" in J. Collins, N. Hall. and L. Paul (eds.), *Causation and Counterfactuals*. Cambridge, Mass.: MIT Press: pp.139-180.

Menzies, P. 2005 "Causal Models, Token-Causation, and Processes", forthcoming in *Philosophy of Science*, Invited Papers, Proceedings of PSA2002.

Pearl, J. 2000. *Causality*. Cambridge: Cambridge University Press.

Russell, B. 1917. "On the Notion of Cause", *Proceedings of the Aristotelian Society*, 13:1-26.

Salmon, W. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

Salmon, W. 1998. *Causality and Explanation*. New York: Oxford University Press.

Sober, E. 1980. "Evolution, Population Thinking, and Essentialism", *Philosophy of Science*. 47: 350-383.

Sober, E. 1984. *The Nature of Selection*. Cambridge: MIT Press.

Unger, P. 1975. *Ignorance: A Case for Skepticism*. Oxford: Oxford University Press.

Van Fraassen, B. The Scientific Image. Oxford: Oxford University Press.