DIFFERENCE-MAKING IN CONTEXT[1]

Peter Menzies

1. Introduction

Several different approaches to the conceptual analysis of causation are guided by the idea that a cause is something that makes a difference to its effects. These approaches seek to elucidate the concept of causation by explicating the concept of a difference-maker in terms of better-understood concepts. There is no better example of such an approach than David Lewis' analysis of causation, in which he seeks to explain the concept of a difference-maker in counterfactual terms. Lewis introduced his counterfactual theory of causation with these words: 'We think of a cause as something that makes a difference, and the difference it makes must be a difference from what would have happened without it. Had it been absent, its effects—some of them, at least, and usually all—would have been absent as well.' (Lewis 1973b: pp. 160-1) According to Lewis, a cause c makes a difference to an effect e in the sense that if the cause c had not occurred, the effect e would not have occurred either. All we shall see in section 2, Lewis' theory says there is more to the concept of causation than this counterfactual condition.

Lewis is on the right track, I think, in saying that we think of a cause as something that makes a difference and that this thought is best explicated in terms of counterfactual concepts. However, I shall argue that the particular way in which Lewis spells out the concept of a cause as difference-maker is unsatisfactory. For Lewis' articulation of this concept is distorted by a specific metaphysical assumption: specifically, that causation is an absolute relation, specifiable independently of any contextual factors.

The distortion induced by this assumption is reflected in the undiscriminating manner in which his theory generates countless causes for any given effect. However, commonsense judgement is much more discriminating about causes than Lewis' theory. Accordingly, I claim that Lewis' analysis faces *the problem of profligate causes* and I outline some specific problem cases in section 3. In the following section I argue that Lewis' most recent formulation of his counterfactual analysis (Lewis 1999a) faces the same problem of profligate causes and also argue that an initially promising solution to the problem that appeals to pragmatics does not succeed.

The key to solving the problem of profligate causation, I argue, is to give up the metaphysical assumption that causation is an absolute relation, specifiable independently of context. In sections 5 and 6 I attempt to analyse the concept of a cause as a difference-maker in a way that integrates a certain contextual parameter into the relevant truth conditions. The analysis employs counterfactual concepts, but ones that are sensitive to context. I use this account in section 7 to explain the problem cases of profligate causation, cited in section 3.

I need to note two restrictions that I intend to impose on my discussion. The first restriction is that I shall only consider cases of deterministic causation. I shall ignore cases of probabilistic causation, not because they do not exist, but

because they do not raise any special issues in connection with the problem of profligate causes.

The second restriction is that I shall only consider cases of non-redundant causation. Any counterfactual rendering of the idea of a cause as a difference-maker must address some tricky questions in dealing with redundant causation—both symmetrical cases involving overdetermination by two or more genuine causes and asymmetrical cases involving pre-emption by two or more potential causes only one of which is genuine. Such examples raise serious questions about the viability of purely counterfactual analyses of causation. For they *seem* to show that a cause need not make a counterfactual difference to its effect due to the presence of alternative causes waiting in the wings. It would take us a long way from our present concerns to determine whether these examples really show this.[2]

My aim in this paper is quite limited. I intend merely to explicate the idea of one condition making a difference to another. I shall claim that the condition expressing this idea is a necessary, though not sufficient, condition for causation. To bring the necessary condition up to sufficiency further concepts have to be added: specifically, I would argue, the concept of a process linking cause with effect. How such a concept is to be added to the difference-making condition and whether the condition, so supplemented, is adequate to deal with cases of redundant as well as non-redundant causation are questions to be pursued on another occasion. For my immediate purposes, it will suffice to have a necessary condition for causation, with which to rule out the profligate causes generated by Lewis' theory.

2. Lewis' 1973 Theory of Causation

Lewis has presented two counterfactual theories of causation: the original 1973 theory and a later 1999 refinement of the theory.[3] In this section I shall discuss his conceptual analysis of causation in the context of the earlier 1973 version of the theory, deferring consideration of the later version of the theory until section 4.

The way Lewis frames his conceptual analysis is influenced by a number of metaphysical assumptions about the causal relation. One of these assumptions, which I will not contest here, is that causation relates dated, localised events (Lewis 1986: pp. 241-69). He means to include events, in the ordinary sense, that involve changes: explosions, battles, conversations, falls, deaths, and so on. But he also means to include events in a broader sense that do not involve changes: a moving object's continuing to move, the retention of a trace, the presence of copper in a sample, and so on.

Another metaphysical assumption Lewis makes about causation is that it is an absolute relation—absolute in the particular sense that it is not relative to any contextual parameter and so does not vary in nature from one context to another. It is this assumption that I wish to contest here. The assumption is not an explicit feature of his analysis, but is rather a consequence of the way he defines the central concept of causal dependence:

> (1) Where c and e are distinct actual events, e *causally depends* on c if and only if e counterfactually depends on c: ie (i) if c were to occur, e would occur; and (ii) if c were not to occur, e would not occur.

Lewis actually works with a simpler definition. He imposes a Centering Principle on the similarity relation governing counterfactuals to the effect that no world is as similar to the actual world as the actual world is to itself. This principle implies that the counterfactual above 'If c were to occur, e would occur' is automatically true in virtue of the fact that c and e are actual events. So his simpler working definition is:

> (2) Where c and e are distinct actual events, e *causally depends* on c if and only if e would not occur if c were not to occur.

The absolute character of causal dependence does not follow from this definition alone. It is possible, after all, to argue that the counterfactual constructions that define causal dependence are to be understood in a context-dependent way. This move is far from implausible in view of the notorious sensitivity of counterfactual constructions to contextual factors. But, in fact, Lewis argues that the counterfactuals that define causal dependence are to be read as non-backtracking counterfactuals; and he specifies the similarity relation that governs them in terms of a unique, context-invariant set of weights and priorities for comparing different respects of similarity (Lewis 1979). The absolute, invariant character of the concept of causal dependence stems ultimately from the absolute, invariant character of the similarity relation for non-backtracking counterfactuals.

A final metaphysical assumption that Lewis makes about causation is that it is a transitive relation. Causal dependence, as defined in terms of counterfactual dependence, is not transitive. To ensure the transitivity of causation, Lewis defines causation in terms of the ancestral of causal dependence:

> (3) Where c and e are distinct actual events, c *is a cause* of e if and only if there is a chain of stepwise causal dependences between c and e.

By defining causation in terms of the ancestral of causal dependence, Lewis is also able to deal with some examples of pre-emption—the so-called examples of early pre-emption. (See Lewis 1986: pp. 193-212.) Though I believe this assumption is false and can be shown to be unnecessary for the treatment of pre-emption examples, I shall not dispute it here.

My present target, to repeat, is the second assumption to the effect that causation is an absolute relation; or more precisely, the assumption that the truth conditions for causal claims can be specified without reference to any contextual factors. Lewis acknowledges that our causal talk is often selective, focusing on some salient causes while backgrounding other less salient causes. However, this selectivity is to be explained by pragmatic principles of

conversational exchange, which leave the objective truth conditions of causal claims untouched.

On Lewis' view, the causal history of an event is a vast, complicated relational structure: the relata in the structure are events and the relation that structures them is causal dependence. Out of this vast structure, the human mind may selectively focus on some part and call it *the* cause of the given event. Indeed, different minds pursuing different enquiries may focus on different parts of this structure. But the 'principles of invidious selection', as Lewis calls them, by which fragments of a causal history are selected for attention, operate on an already fully determinate causal history. The principles of selection are independent of the relational structure itself. In this connection he writes:

> The multiplicity of causes and the complexity of causal histories are obscured when we speak, as we sometimes do, of *the* cause of something. That suggests that there is only one…. If someone says that the bald tyre was the cause of the crash, another says that the driver's drunkenness was the cause, and still another says that the cause was the bad upbringing that made him reckless, I do not think that any of them disagree with me when I say that the causal history includes all three. They disagree only about which part of the causal history is most salient for the purposes of some particular enquiry. They may be looking for the most remarkable part, the most remediable or blameworthy part, the least obvious of the discoverable parts…. Some parts will be salient in some contexts, others in others. Some will not be salient in any likely context, but they belong to the causal history all the same: the availability of petrol, the birth of the driver's paternal grandmother, the building of the fatal road, the position and velocity of the car a split second before the impact. (Lewis 1986: pp. 215-6)

3. The Problem of Profligate Causation

How satisfactorily does this theory capture the idea of a cause as a difference-maker? A fair answer would have to be: 'Not very well' or 'Well enough, but only with a lot of auxiliary assumptions about pragmatics'. The main problem that I wish to highlight here is the profligate manner in which the theory generates causes for any given effect. Below I list a number of examples that illustrate this defect of the theory. They are familiar examples for the most part, but it is useful to have a catalogue of them before us.

According to Lewis' theory, any event but for which the effect would not have occurred is one of its causes. But, as is widely recognised, this generates some absurd results.

> *Example 1: The Lung Cancer*
> A person develops lung cancer as a result of years of smoking. It is true that if he had not smoked he would not have developed cancer. It is also true that he would not have developed lung cancer if he had not

> possessed lungs, or even if he had not been born. But it is absurd to think his possession of lungs or even his birth caused his lung cancer.

Commonsense draws a crucial distinction between causes and background conditions. It ranks the person's possession of lungs and his birth as background conditions, so disqualifying them from being difference-makers for the effect. Several philosophers of causation have stressed the importance of this commonsense distinction in connection with the view of causes as difference-makers. J. L. Mackie, for example, says that what we call a cause is 'what makes the difference in relation to some assumed background or causal field' (Mackie 1974: p. 35).[4]

Perhaps the most extensive and penetrating investigation of the distinction between causes and conditions is that of H. L. Hart and A. Honoré in their seminal work *Causation in the Law* (1985). They argue that the distinction is relative to context in two different ways. One form of relativity might be called *relativity to the context of occurrence*.[5] They contrast our causal judgements about the following situations.

> *Example 2: The Presence of Oxygen*
> If a building is destroyed by fire, it may be true that the fire would not have taken hold but for the oxygen in the air, the presence of combustible material, and the dryness of the building. But these are mere *conditions* of the fire. On the other hand, if a fire breaks out in a laboratory or in a factory, where special precautions are taken to exclude oxygen during the experiment or manufacturing process, it would not be absurd to cite the presence of oxygen as a *cause* of the fire. In both situations it may be true that the fire would not have occurred if oxygen had not been present. (Modified from Hart and Honoré 1985: pp. 35-6)

The second form of context-relativity might be called *relativity to the context of enquiry*. With this form, rather than two different situations eliciting different judgements about causes and conditions, as in the example above, it is one and the same situation that elicits different judgements depending on the type of enquiry being undertaken. Here are some of Hart and Honoré's examples.

> *Example 3: The Famine and the Ulcerated Stomach*
> 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 reserves as the cause and the drought as a mere condition. Someone who prepares meals for a person suffering from an ulcerated condition of the stomach might identify eating parsnips as the cause of his indigestion, but a doctor might identify the ulcerated condition of the stomach as the cause and the meal as a mere condition. (Modified from Hart and Honoré 1985: pp. 35-6)

Lewis' theory is insensitive to the different context-relative ways in which commonsense draws the distinction between causes and conditions. His theory treats mere conditions as causes because they are factors without which the effect would not have taken place.[6]

Hart and Honoré argue convincingly, in my view, that the suggestions made by various philosophers for drawing the commonsense distinction between causes and conditions are unsatisfactory. They reject J. S. Mill's suggestion that the distinction is the epistemically based distinction between causal factors revealed by investigation and causal factors known before investigation. They point out that we would count a dropped cigarette as the cause of a fire even when we learn from science, what we may not have initially known, that the presence of oxygen is among the conditions required for its occurrence. Hart and Honoré also reject R.G. Collingwood's suggestion that the distinction is the practically based distinction between factors controllable by the investigator and factors not so controllable. They argue that the discovery of the cause of cancer would still be the discovery of the cause, even if it were useless for the cure of the disease; and that drought is the cause of the failure of crops and so of famine, and lightning the cause of a fire even for those who can do nothing to prevent them (Hart and Honoré 1985: pp. 34-7).

Hart and Honoré also argue that it is wrong to identify the conditions as the ordinary course of nature unaffected by human intervention. They observe that the commonsense distinction is very often an artefact of human habit, custom or convention. Because nature can be harmful unless we intervene, we have developed customary techniques, procedures and routines to counteract such harm. These become a second 'nature'. For example, the effect of a drought is regularly neutralised by government precautions in conserving water; disease is neutralised by inoculation; rain by the use of umbrellas. When such procedures are established, the cause of some harm is often identified as an omission or failure on the part of some agent to carry out the neutralising procedures, as the example of the famine illustrates (Hart and Honoré 1985: pp. 37-8).

The fact that omissions, absences, and failures are recognised as causes and effects poses a *prima facie* difficulty for Lewis' theory, which requires causation to link events. Lewis concedes that an absence is a bogus kind of entity that cannot be counted as an event. Nonetheless, in partial solution to this difficulty, he argues that the proposition that an absence occurs is not bogus; and since such propositions can enter into counterfactual dependences, we can talk in a derivative way about causation by absences.

Lewis' extension of causal relata to include absences exacerbates the problems already noted. With the inclusion of absences as possible causes of a given effect, the blurring of the distinction between causes and conditions by Lewis' theory generates even more counterintuitive results. The next examples are illustrative of the difficulties Lewis' theory encounters with absences and other non-occurrences.

Example 4: The Absence of Nerve Gas
I am writing this paper at my computer. If, however, there were nerve gas in the air, or I were attacked with flamethrowers, or struck by meteor shower, I would not be writing the paper. But it is counterintuitive to say that the absence of nerve gas, flamethrower attack, and meteor strike are causes of my writing the paper.

Example 5: The Multiple Omissions
A healthy plant requires regular watering during sweltering hot weather. A gardener whose job it is to water the plant fails to do so and the plant dies. But for the gardener's omission, the plant would not have died. On the other hand, if anyone else had watered the plant, it would not have died. But it seems to absurd to say that the omission of everyone else to water the plant was a cause of its death.

Commonsense draws a distinction among these negative occurrences, ranking some of them as causes and others as mere conditions. In the first example, it rates all the absences as conditions, but in the second example it distinguishes the gardener's omission from the other omissions, making it a cause. In contrast, Lewis' theory treats all these non-occurrences equally as causes.

In summary, the examples cited in this section all point to the counterintuitive causal judgements licensed by Lewis' theory. They illustrate that his theory, or at least its truth conditional component, is too profligate in its attribution of causes because it does not respect the context-relative way in which commonsense distinguishes between causes and conditions.

4. A Possible Defence in Terms of Contrastive Explanation

We have so far been considering the original 1973 version of Lewis' counterfactual theory. Does the more recent 1999 version of the theory fare any better with these problematic examples?

The more recent version of the theory also employs a counterfactual rendering of the idea that a cause makes a difference to its effect. But the counterfactuals it employs do not simply state dependences of *whether* one event occurs on *whether* another event occurs. The counterfactuals state dependences of *whether*, *when*, and *how* one event occurs on *whether*, *when*, and *how* another event occurs. A key idea in the formulation of these counterfactuals is that of an *alteration of an event*. This is an actualised or unactualised event that occurs at a slightly different time or in a slightly different manner from the given event. An alteration is, by definition, a very fragile event that could not occur at a different time or in a different manner without being a different event. Lewis stipulates that one alteration of an event is the very fragile version that actually occurs.

The central notion of the new version of the theory is that of influence:

(4) If $\underline{c}$ and $\underline{e}$ are distinct events, $\underline{c}$ *influences* $\underline{e}$ if and only if there is a substantial range of $\underline{c}_1$, $\underline{c}_2$,... of different not-too-distant alterations of $\underline{c}$

> (including the actual alteration of $\underline{c}$) and there is a range of $\underline{e}_1$, $\underline{e}_2$,... of alterations of $\underline{e}$, at least some of which differ, such that if $\underline{c}_1$ had occurred, $\underline{e}_1$ would have occurred, and if $\underline{c}_2$ had occurred, $\underline{e}_2$ would have occurred, and so on.

Where one event influences another, there is a pattern of counterfactual dependence of whether, when, and how upon whether, when, and how. As before, the notion of causation is defined as an ancestral relation.

> (5) $\underline{c}$ *causes* $\underline{e}$ if and only if there is a chain of stepwise influence from $\underline{c}$ to $\underline{e}$.

This theory is designed to handle cases of redundant causation. For example, consider a standard case of late pre-emption. Billy and Suzy throw rocks at a bottle. Suzy throws first so that her rock arrives first and shatters the bottle. However, without Suzy's throw, Billy's throw would certainly have shattered the glass. Suzy's throw is the pre-empting cause of the shattered bottle, Billy's throw the pre-empted potential cause. The revamped theory explains why we take Suzy's throw, and not Billy's throw, to be the cause of the shattering of the bottle. If we take an alteration in which Suzy's throw is slightly different—the rock is lighter or she throws sooner—(while holding Billy's throw fixed), we find that the shattering is different too. But if we make the same alterations to Billy's throw (while holding Suzy's fixed), we find that the shattering is unchanged. (Lewis 1999a: p. 24)

The important question for our purposes is whether this new version of the theory helps to deal with any of the problematic examples above. I cannot see that it helps with any of them. The theory might try to explain the distinction between causes and conditions by showing that effects are sensitive to alterations in causes in a way that they are not sensitive to alterations in conditions. But this does not seem to be the case. In the Lung Cancer example, for instance, altering the man's possession of lungs will change the way his lung cancer develops as surely as altering his habits of smoking. Effects are sensitive to alterations in conditions as much as to alterations in causes. So the theory does not do anything to explain the distinction between causes and conditions.

Indeed, Lewis concedes as much himself. He says that the new version of the theory, as much as the old version, appears to generate too many causes. Any presence or absence linked by a pattern of influence, or a chain of such patterns, will count as a cause, although this seems to go against our ways of thinking and speaking of causes. However, he offers a defence of his theory in terms of Grice's (1975) pragmatic theory of conversational implicature. It is literally true that any presence or absence linked to an event in the way distinguished by his theory is a cause of that event, but it is not always conversationally appropriate to mention it as a cause. He writes: 'There are ever so many reasons why it might be inappropriate to say something true. It

might be irrelevant to the conversation, it might convey a false hint, it might be known already to all concerned….' (Lewis 1999a: p.34)

Lewis belongs to a long tradition of philosophers who have tried to isolate objective truth conditions for causal statements from pragmatic considerations of context. J. S. Mill famously claimed that the only objective sense of cause is that of the total cause of some effect. He dismissed the context-dependent way in which we ordinarily talk of some partial conditions as causes and others as conditions.[7] Of course, Lewis' position is slightly more subtle than Mill's. Whereas Mill dismissed our ordinary talk as unsystematic and muddled, Lewis gestures at the outlines of a possible explanation in the form of Grice's maxims of conversational exchange.

However, Lewis provides scant detail of the way Grice's maxims are meant to apply to particular examples. Which maxims are relevant? How are they to be employed? There is, moreover, a question whether Grice's principles are especially well suited to explaining the specific causal judgements in question. Grice's maxims of conversation are very general principles of rationality applied to information exchange. Yet the principles that lie behind our judgements about the examples of the last section seem to be particular to causal judgements. As general principles of rational information exchange, Grice's maxims miss out on these particular causation-specific principles.

What are these causation-specific principles? Several philosophers investigating the pragmatics of causal explanation have stressed the importance of using contrastive why-questions to analyse the interest-relativity of causal explanation. (See, in particular, Garfinkel 1981 and van Fraassen 1980: Chapter 5.) They have argued that seeing causal explanations as answers to contrastive why-questions affords a way of understanding how context filters out from the vast causal history of an event those causes that are salient for certain explanatory concerns. For example, an enormous range of causal factors can be cited in explanation of a particular eclipse of the moon. Nonetheless, we can restrict attention to certain kinds of causal factors by seeking explanations of specific contrasts: Why did the eclipse occur rather than not occur? Why was it partial rather than complete? Why did it last two hours rather than some other interval of time? Different contexts can be seen as implicitly requesting explanations of these different contrasts. This work on contrastive explanation suggests a strategy for explaining our causal judgements about the examples of the last section: preserve Lewis' counterfactual analysis of causation, but add on to it an account of contrastive explanation that can explain the context-sensitivity of ordinary causal discourse.[8]

There are several accounts of contrastive explanation available. Which should we use? Lewis has developed one account, which, as it happens, is tailormade for our purposes, though it must be noted that Lewis himself does not envisage the applications to which we shall put it. He writes in his paper 'Causal Explanation':

> One way to indicate what sort of explanatory information is wanted is through the use of contrastive why-questions. Sometimes there is an

> explicit 'rather than...'. Then what is wanted is information about the causal history of the explanandum event, not including information that would also have applied to the causal histories of alternative events, of the sorts indicated, if one of them had taken place instead. In other words, information is requested about the difference between the actualised causal history of the explanandum and the unactualised causal histories of its unactualised alternatives. Why did I visit Melbourne in 1979, rather than Oxford or Uppsala or Wellington? Because Monash invited me. That is part of the causal history of my visiting Melbourne; and if I had gone to one of the other places instead, presumably that would not have been part of the causal history of my going there. It would have been wrong to answer: Because I like going to places with good friends, good philosophy, cool weather, and plenty of trains. That liking is also part of the causal history of my visiting Melbourne, but it would equally have been part of the causal history of my visiting any of the other places, had I done so. (Lewis 1986: pp.229-30)

On this account, we explain why an event e rather than an event e* occurred by giving information about the actual causal history of e that differentiates it from the counterfactual causal history of e*.

Can the strategy of conjoining this account with the counterfactual analysis help to explain our intuitive judgements about the examples of section 3? Let us be clear about what the strategy is in the first place. It involves two auxiliary assumptions. The first is the assumption that every ordinary causal statement can be seen as a response to an implicit contrastive why-question. The second is the assumption that our ordinary talk about causes is to be explained in terms of the way contrastive why-questions selectively filter out from the objective causes, delivered by the counterfactual analysis, those relevant in particular contexts.

Let us consider how well this strategy works by seeing how it applies to the examples of section 3. It must be said that it works surprisingly well with some of them. For example, it explains reasonably well why we do not consider it appropriate to say about the Lung Cancer example that the man's birth and his possession of lungs were causes of his lung cancer. For it is natural to assume that such causal statements would be attempts to answer the why-question 'Why did the man get lung cancer rather than not?'. However, the man's birth and possession of lungs fail to be objective causes that are present in the actual causal history of his lung cancer but absent from the counterfactual causal history of his not getting lung cancer. If the man had not developed lung cancer, it would still plausibly be part of his causal history that he was born and possessed lungs.

The strategy also provides a convincing explanation of the relativity of our causal judgements about causes and conditions to the context of enquiry. For instance, in the Ulcerated Stomach example, the meal-preparer and the doctor make different judgements about causes and conditions because they address different contrastive why-questions. The meal-preparer is addressing

the question ‘Why did the person get indigestion on *this occasion* rather than some other?' The person’s ulcerated stomach is a condition that is present in both the actual history and the counterfactual histories, and so disqualified from counting as a factor that differentiates between them. On the other hand, the doctor is addressing the question ‘Why does *this person* rather than other people get indigestion?’. Here what the person ate is a factor common to his causal history and the causal histories of other people, while his ulcerated stomach condition is a factor that differentiates them.

Without doubt, these explanations of our commonsense judgements have a ring of plausibility to them. Nonetheless, I think they cannot be the complete story, as the principles they rely on have some major gaps or inadequacies.

First, Lewis’ account of contrastive explanation relies on backtracking counterfactuals. We have to be able to work out whether some objective cause would be present or absent from the history that would have had to occur if some alternative to the actual effect had occurred. But the principles that guide the reasoning behind such backtracking counterfactuals have not been formulated. For example, to get the right results in the Lung Cancer example we have to infer that that the person would have been born and possessed lungs even if he had not developed lung cancer. And to get the correct answer in the Absence of Nerve Gas example we have to infer that if I were not writing at my computer, it would not be because nerve gas had been intruded into my office, or because I had been attacked by flamethrowers, or been struck by a meteor. But why are these inferences alone reasonable? Backtracking reasoning, unguided by any principles or unconstrained in any way, could equally well lead to the opposite conclusions. Clearly, this strategy, if it is to give the correct verdicts about the examples, must articulate some fairly detailed principles regarding the appropriate kind of backtracking reasoning. Until these principles have been articulated, the strategy is incomplete.

Secondly, one of the central assumptions of the strategy—namely, that every causal statement must be understood in the context of an implicit contrastive why-question— is too strong. This assumption may hold for some cases, but it is dubious whether it holds for all. This becomes clear if we allow, as I think we should, for cases of probabilistic causation, in which the cause brings about the effect but does so with a chance of less than one. It may be true, for instance, that bombarding a radioactive particle causes it to decay, but the bombardment is not something that differentiates the actual causal history leading to decay from the counterfactual history leading to non-decay. For the atom may fail to decay in the counterfactual history, not because the atom is not bombarded, but because the bombardment does not, as a matter of pure chance, lead to decay.

Thirdly, Lewis’ account of contrastive explanation does not capture an important feature of contrastive explanations. This point is clearer where contrasts between compatible alternatives, rather than incompatible alternatives, are being explained. An example of a contrast between compatible alternatives is Carl Hempel’s (1965) much discussed example of syphilis and

paresis. Paresis is a late developmental stage of the disease syphilis, but, as it happens, few people with syphilis contract paresis. Nonetheless, we can still explain why Jones, rather than Smith, contracted paresis by saying that only Jones had syphilis. But this cannot be the right explanation on Lewis' account: for syphilis does not differentiate between the actual case in which Jones gets paresis and the counterfactual history in which Smith gets paresis, since the only way in which Smith could get paresis is by first developing syphilis. Such examples highlight a feature of contrastive explanations not captured in Lewis' account. Sometimes the correct contrastive explanation compares actual with actual, rather than actual with counterfactual. In the example under consideration, it cites an actual feature that differentiates Smith and Jones—a feature present in Jones' case but absent from Smith's case.

Finally, and most importantly, the two-part strategy is very unsatisfactory from an explanatory point of view. For it unnecessarily duplicates the use of the idea of a cause as something that makes a difference: first in the analysis of 'objective cause' as something that makes a counterfactual difference; and then again in the contrastive explanation account of the 'context-sensitive cause' as something that differentiates actual from counterfactual histories. These uses of the idea are clearly independent, with neither being derived from the other. Yet it would surely be a surprising fact, requiring elaborate explanation, if our framework for conceptualising causation used in two different but crucial ways the very same idea of difference-making. It would be much more likely that our conceptual framework developed on the basis of a single fundamental application of this idea.

For these reasons, then, the two-part strategy is not as promising as it first appeared. What is really required is a unitary account of causes as difference-makers that explains the success of this strategy while avoiding its failures. In my view, if we are to develop such an account, we must draw a distinction between two kinds of theories of the context-sensitivity of causal discourse. *Add-on context-sensitive theories*, like Lewis', apply pragmatic principles such as Grice's maxims or principles about contrastive explanation to independently determined truth conditions. In contrast, *integrated context-sensitive theories* make the context-sensitivity intrinsic to the truth conditions of causal claims by making the truth conditions relative to certain contextual parameters. I shall recommend adopting an integrated rather an add-on account of causal claims.[9]

5. Causal Models

As we have seen, Lewis' view that causation relates events is confounded by the fact that commonsense also allows absences and omissions as causes and effects. The difficulty shows that we need to be inclusive about the relata of causation. In order to be as inclusive as possible, I shall talk of *factors* as causes and effects. Factors are meant to include anything that commonsense dignifies as causes and effects—events, states of affairs, absences, omissions, and other non-occurrences.[10] I shall reserve the uppercase variables C, D, E and so on for factors.

Any theory of causes as difference-makers must make a connection with Mill's Method of Difference for detecting causes and testing causal claims. A crucial part of the method is a *difference observation* between a positive instance in which some effect E is present and a negative instance in which E is absent. If some condition C is present in the positive instance and absent in the negative instance, it is, at least, part of what makes the difference to E. Mackie (1974: pp. 71-2) points out that there are two forms that the classical difference observation can take. One form is the before-and-after experiment in which some change C is introduced, either naturally or by deliberate human action, into an otherwise apparently static situation . The state of affairs just after the introduction is the positive instance and the state of affairs just before it is the negative instance. If the introduction is followed, without any further intervention, by some change C, then we reason that C is part of what made the difference to E. The other form the classical difference observation can take is the standard controlled experiment, where what happens in the experimental case is compared with what happens in a deliberately controlled case which is made to match the experimental case in all ways thought likely to be relevant other than C, whose effects are under investigation.

Mackie points out that different conceptual analyses of causes as difference-makers are modelled on the two forms of the classical difference observation. For example, C. J. Ducasse's (1968) theory of causation is clearly modelled on the before-and-after observation. It states that the cause of a particular change E is the particular change C that alone occurred in the close environment of E immediately before it. However, I agree with Mackie that this analysis is inadequate as an account of causation, as it fails to distinguish between causal and non-causal sequences of events. Consider Mackie's pair of contrasting sequences (1974: p. 29). In one sequence, a chestnut is stationary on a flat stone. A person swings a hammer down so that it strikes the chestnut directly from above and the chestnut is flattened. In the other sequence, a chestnut is stationary on a hot sheet of iron. A person swings a hammer down so that it strikes the chestnut directly from above. At the very instant the hammer touches it, the chestnut explodes with a loud pop and its fragments are scattered around. Couched as it is in terms of actual changes, Ducasse's theory is hard pressed to deliver the correct verdict that the hammer blow is a cause in the first sequence but not in the second.

Mackie argues that these examples show that the relevant contrast in the difference observation is not the before-and-after contrast, but the experimental-and-control contrast (1974: Chapter 2). We judge that the hammer blow is the cause of the effect in the first sequence because if we were to intervene in the course of events to prevent the hammer from striking the chestnut, the flattening would not occur; and we judge that the hammer blow is not the cause of the effect in the second sequence because if we were to intervene to prevent the hammer striking the chestnut, the explosion would still occur. Mackie argues that the conceptual analysis based on the experimental-and-control form of the difference observation must appeal to modal notions, in particular to conditionals. More specifically, he argues that

the conceptual analysis of cause as difference-maker must appeal to two conditionals, one factual and the other counterfactual:

> (6) Where C and E are distinct factors, C makes a difference to E if and only if E would occur if C were to occur and E would not occur if C were not to occur.

This analysis captures the idea involved in the experimental-and-control contrast precisely because one conditional represents what happens in the experimental case and the other what happens in the control case.

I find much of what Mackie says about the experimental-and-control contrast idea very illuminating. However, his discussion of this idea is marred by confusions about the conditionals that are supposed to capture this contrast. Especially confusing is his meta-linguistic account of conditionals, according to which they do not have truth conditions. Nonetheless, I am going to take, as a starting point for my discussion, Mackie's claim that the experimental-and-control form of the difference observation is the relevant analogical basis for a conceptual analysis of difference-making. I shall also take, as a starting point for my discussion, the thesis that this contrast can be spelled out in terms of a pair of conditionals, one representing what happens in the experimental case and the other representing what happens in the control case. (Unfortunately, we shall have to wait until the next section to see the full justification for these assumptions.) However, I shall reject Mackie's confusing account of conditionals, in favour of more orthodox truth conditional account in terms of possible worlds. Under such an account, the central idea of difference-making can be spelled out in the following schematic terms.

> (7) Where C and E are distinct factors, C *makes a difference* to E if and only if every most similar C-world is an E-world and every most similar ~C-world is a ~E-world.

This formulation neatly captures the idea of a cause as a difference-maker: where two relevantly similar possible worlds differ with respect to C they also differ with respect to E, and vice versa. A condition that just happens to covary with another in their actual instances will not modally covary in the way required to count as making the difference.

I should state at the outset that, while using the standard possible worlds framework for understanding conditionals, I understand the possible worlds in a slightly unconventional way. The possible worlds I shall employ are mini-worlds rather than alternative large-scale universes: they are alternative courses of development of typically small-scale systems. They are best understood as being similar to the trajectories in the state space posited by a scientific theory to describe the behaviour of systems of a certain kind. A theory may seek to describe the behaviour of a certain kind of system in terms of a set of state variables $\{S_1,..., S_n\}$. The accompanying state space will be an n-dimensional space and the trajectories in this space will be temporally ordered

sequences of states in this space. So while I use the traditional term ‘possible world’, it should always be kept in mind that I understand it typically in the ‘mini-world’ sense, where the mini-worlds are understood as analogous to trajectories in a state space for a typically small-scale system of a certain kind.

The all-important question to be answered about the possible worlds formulation of the difference-making idea is: Which worlds count as relevantly similar to the actual world? As we have seen, Lewis thinks that, for each causal claim about an event that makes a difference to another, the corresponding counterfactuals are to be read in terms of a unique kind of similarity relation. In this respect, my position differs from Lewis’, in that I think that the relevant similarity relations are context-dependent, with causal statements in different contexts requiring different similarity relations. Causal statements must be understood, I shall argue, as relative to a certain contextual parameter; and depending on the way the parameter is set, an appropriate kind of similarity is determined for a given causal statement.

The contextual parameter in question reflects the fact that our causal thinking is steeped in abstraction. It is a platitude—but one worth repeating—that the world is exceedingly complex in its causal structure. Within any spatiotemporal region, there are many different levels of causation, and within each level many crosscutting and intersecting causal processes. In order to determine the structure of these processes, we are necessarily forced, by the finitude of our minds, to focus selectively on some aspects of what is going on and to ignore or background others. The causal schemas by which we interpret the world are irremediably permeated by abstractions that enable this selective focusing. There seem to be several forms of abstraction that underlie our causal thinking.

One form of abstraction underlying our thinking about the causal structures of a concrete situation involves the identification within the situation of *a particular set of objects as forming a system of a certain kind*. A particular system may consist of a great many objects or very few, of very large objects or very small ones. Astronomers and cosmologists investigate vast systems—solar systems, galaxies, or the whole cosmos. The systems investigated by biologists and economists—economies, markets, species, populations, and so on—are smaller, but still large by human standards. On the other hand, the systems investigated by particle physicists are small by any standard. It is not always easy to determine which objects belong to a particular system. This is not just because of our epistemic limitations, but because the spatiotemporal boundaries of the system are indeterminate. How many astronomical bodies are in the Milky Way Galaxy? How many organisms belong to a population of marsh frogs? It is difficult to answer these questions because the spatiotemporal boundaries of these systems are not perfectly determinate. Nonetheless, the indeterminate localisation of systems does not stop scientists from conceptualising causal structures in terms of them.

The form of conceptual abstraction under consideration involves not just the identification of a particular set of objects, but the identification of this set of objects as constituting *a system of a certain kind*. But what is a system? A

simple answer to this question is that a particular system is a set of objects that have certain properties and relations. But not any old properties and relations are relevant to the identification of a system. For example, a set of astronomical bodies can be individuated as a particular planetary system by way of each astronomical body's relation to other bodies in the system, but not by way of their relations to objects outside the system; a particular population of marsh frogs may be individuated in terms of the frogs' relational property of living in a particular marsh, but not in terms of extraneous relational properties involving far-distant objects. In short, a system is a set of constituent objects that is *internally organised* in a distinctive fashion; and the properties and relations that configure the objects into a system must be *intrinsic* to the set of constituent objects.

The concept of intrinsic properties and relations has been much discussed. However, the significant concept under consideration here is not the concept of properties and relations that are intrinsic *tout court*, but those that are intrinsic to a set of objects. It will suffice for our purposes to explain the intuitive idea behind these concepts, rather than to present a full analysis of them, which turns out to be slightly tricky. Modifying an idea of Jaegwon Kim's (1982) concerning the simple concepts, I shall say that:

> (8) A property F is *extrinsic to a set of objects* if and only if necessarily, one of its members has F only if some contingent object wholly distinct from the set exists.

For example, the extrinsic properties of a set of astronomical bodies would include being observed by some human and being a certain distance from the earth (assuming the earth is not in the set).[11]

The concept of a property intrinsic to a set of objects is defined in converse fashion:

> (9) A property F is *intrinsic to a set of objects* if and only if possibly, one of its members has F although no contingent object wholly distinct from the set exists.

For example, the intrinsic properties of a set of astronomical bodies would include the mass and shape of the individual astronomical bodies. But the intrinsic properties of the set need not be all intrinsic properties *simpliciter*. For example, the property of being gravitationally attracted to another body that is also a member of the set is an intrinsic property of the set, though it is not an intrinsic property *simpliciter*.[12]

There are, literally, uncountably many particular systems, but very few of them are of any interest to us. For the most part, we are interested in the *kinds of systems* that evolve in lawful ways. For example, certain systems of astronomical bodies and certain systems of biological organisms have intrinsic properties and relations which change over time in regular ways described by certain laws. Identifying a kind of system involves identifying the intrinsic

properties and relations that are shared by particular systems and that conform to certain laws. The state variables employed in a scientific theory correspond to the intrinsic properties and relations that constitute a kind of system. In Newtonian mechanics, for instance, the state variables used to describe the behaviour of mechanical bodies are the properties of mass, position, and momentum. The following definition captures these ideas:

> (10) *A kind of system* K is a set of particular systems sharing the same intrinsic properties and relations (state variables) whose evolution over time conforms to certain laws.

By definition, the state variables that determine a given kind of system are intrinsic properties and relations of the particular systems belonging to the kind. More generally, a kind of system supervenes on a set of intrinsic properties and relations in the sense that any two particular systems with the same intrinsic properties and relations must both belong, or fail to belong, to a given kind of system.

I have said that a certain contextual parameter determines the similarity relation relevant to working out whether some condition makes a difference to another in a given concrete situation. I propose that one element of this contextual parameter is a kind of system. It is, I claim, an automatic and inevitable feature of the way in which we conceptualise the causal relations of a concrete situation that we see the concrete situation as an instance of a certain kind of system.

The other element in the similarity-determining contextual parameter is the set of laws governing the kind of system under consideration. This element of the contextual parameter reflects a further type of abstraction involved in our causal thinking. For almost invariably the laws governing the kinds of system of interest to us are *ceteris paribus* laws. Such laws state that the relevant systems evolve along certain trajectories provided nothing interferes. For example, the law of gravity in Newtonian mechanics states that, *provided there is no other interfering force*, the force exerted by one object on another varies directly as the product of their masses and inversely as the square of the distance between them. The law of natural selection states that, *provided there is no force besides that of selection at work*, if organisms possessing a heritable trait F are fitter than organisms with an alternative heritable trait F', then the proportion of organisms in the population having F will increase. Geoffrey Joseph (1980) suggests that such laws would be better called *ceteris absentibus* laws, as they usually describe the evolution of the relevant systems under the assumption that all interfering factors or forces are *absent*. Such an assumption is, often enough, an idealisation, because most kinds of systems are subject to interfering influences in addition to the causal influences described by their relevant laws.

Idealisation is central to our causal thinking, as is evident from the ubiquity of *ceteris paribus* laws. Still, many philosophers have thought that *ceteris paribus* laws are disreputable in some way. For example, it is sometimes

objected that *ceteris paribus* laws are vacuous because the *ceteris paribus* condition cannot be specified non-trivially. (See, for instance, Fodor 1991.)The law that *ceteris paribus* all Fs are Gs, the objection runs, is really just the vacuous law that all Fs, unless they are not Gs, are Gs. This objection has no cogency at all, in my view. The law of gravity that tells us how, in the absence of other causal influences, gravity exerts a force on objects is far from trivial. It makes a substantive claim about the world because the concept of an interfering causal influence can be explicated informatively. Without being overly precise, we can explicate the concept in the following terms:

> (11) A factor I is *an interfering factor* in the evolution of a system of kind K in conformity with the laws L if and only if:
> (i) I instantiates an intrinsic property or relation in a particular system of kind K;
> (ii) I is caused by some factor instantiating a property or relation extrinsic to the system of kind K; and
> (iii) the laws governing the causation of I by the extrinsic factor are distinct from the laws L.

Condition (i) simply states that the interfering factor is an intrinsic feature of the system in question. But condition (ii) says that this factor must have a causal source extrinsic to the system. Condition (iii) says that the causation of the factor can be explained independently of the laws governing the system in question. The paradigm example of an interfering factor is the result of an intervention by a human agent in the workings of the system. For example, the gravitational force of the earth on a simple pendulum can be counteracted by a simple human intervention in the swing of the pendulum. While human interventions are not the only kinds of interfering factors, they form the analogical basis for our thinking about interfering forces. For they constitute the most familiar type of situation in which an external force, operating according to its own distinctive laws, can intervene or intrude into the workings of a system.

Given this explication, we can see that the hypothesis that the *ceteris paribus* laws of Newtonian mechanics hold true of some system, say the system of planets orbiting around the sun, involves a substantial claim about the world. For the hypothesis commits one, not only to making certain predictions about the orbits of the planets, but also to explaining prediction failures in terms of the external interfering forces whose causal explanation is, in some sense, independent of the system and laws under consideration. *Ceteris paribus* laws are, to use the words of Pietroski and Rey (1995), like ‘cheques’ written on the banks of independent explanations, their substance and warrant deriving from the substance and warrant of those explanations. It may be questionable on some occasions whether the cheque can be cashed, but that hardly demonstrates the general inadequacy of the institution of bank cheques.

Another common objection to *ceteris paribus* laws is that, even if *ceteris paribus* clauses can be specified non-trivially, they cannot be specified

determinately. (See, for instance, Schiffer 1991.) For it is impossible to specify in advance all the interfering factors whose absence is required to enable a given system to evolve in accordance with given laws. Without doubt, there is truth in this claim. But it is a mistake to think this somehow impugns the determinacy of a *ceteris paribus* law. It is a mistake to say that the statement 'There is only one person in the room'—alternatively 'There is one person in the room and no one else'—has no determinate sense because one cannot specify in advance every person whose absence is required to verify the negative existential. This mistake rests on a confusion about what the determinacy of negative existentials requires. The objection to the determinacy of *ceteris paribus* laws rests on exactly the same confusion.

To capture the fact that our causal thinking is permeated by the two kinds of abstraction identified above, I shall say our causal judgements about a concrete situation must be understood as relative to a *causal model* of the situation. I represent a causal model of a situation as an ordered pair <K, L>, where the first element K is the kind of system in terms of which we conceptualise the situation, and the second element L is the set of laws, typically *ceteris paribus laws*, governing the evolution over time of that kind of system. In using the term 'causal model', I hope to highlight the continuity between commonsense causal thinking and the causal theorising of the natural and social sciences. Several philosophers of science—notably, Nancy Cartwright (1983, 1999), Ronald Giere (1988), Fred Suppe (1979, 1989), and Bas van Fraassen (1980, 1989)—have emphasised that theorising in these sciences often proceeds by way of idealised causal models in which *ceteris paribus* laws play a central, indispensable role. I would claim that these features of scientific practice have their roots in everyday causal reasoning.

No doubt the claim that causal judgements about a concrete situation are to be understood as relative to a causal model of the situation will strike many as confused and erroneous. So let me try to forestall some misunderstandings of this claim.

First, I am *not* claiming that causation is mind-dependent in some idealist sense. I am simply explicating the scientific commonplace that the causal structure of any particular situation can be modelled in several different ways. I interpret this commonplace as meaning that a given situation can be viewed as instantiating different kinds of systems obeying different laws. In the analysis to follow, the claim that the difference-making relation is relative to a causal model M = <K, L> should be understood in terms of a conditional construction of the following form: *if* the given situation instantiates a kind of system K governed by laws L, *then* C makes a difference to E if and only if...', where what replaces the dots will state a perfectly objective condition about the world. If the given situation satisfies the antecedent of this conditional, it is a completely mind-independent matter whether some factor in the situation makes a difference to another.

Secondly, I am *not* endorsing a crude relativism to the effect that any causal model of a situation is as good as any other, or more especially, any kind of system is just as natural as any other for determining causal relations. There

are natural kinds, in my view, but it is the job of metaphysics and science rather than conceptual analysis to investigate what they are. However these investigations turn out, a plausible metaphysics is likely to allow that any particular spatiotemporal region instantiates several different kinds of systems. Perhaps an extremely austere physicalism committed to the existence of a unified field theory would assert that every situation is to be modelled in terms of a unique physical kind of system subject to the unified field equations. However, any less austere metaphysics is likely to conclude that several, perhaps imperfectly natural, kinds of systems may be instantiated in a given spatiotemporal region. In this case, a conceptual analysis should be able to make sense of the alternative causal judgements about these different kinds of systems.

Finally, I am *not* saying that a causal model must be specified in terms of known kinds of systems and known laws. My discussion has been influenced by philosophers of science who argue that scientific theories are best understood as abstract models. Of practical necessity, they discuss known scientific theories in terms of known kinds of systems and known laws. But I do not wish to confine causal models to what is actually known. Commonsense and scientific practice accept a realism that states that we may be ignorant of the intrinsic properties and relations that constitute a kind of system, and we may yet have to discover all the laws governing a kind of system. Indeed, our causal judgements may presuppose a causal model that can be specified imperfectly only in terms of an incompletely known kind of system and set of laws. But the analysis to follow can proceed satisfactorily in terms of an *objectified* causal model along these lines: if the given concrete situation is an instance of this imperfectly known kind K obeying the imperfectly known laws L, then a difference making claim is true if and only if....

6. The Similarity Relation and Difference-making

How exactly are causal judgements about a concrete situation relative to a causal model? The relativity of causal judgements to a model consists, I shall argue, in the fact that the model determines the respects of similarity used in evaluating whether a putative cause makes a difference to an effect. I will try to explain the way a model determines these respects of similarity in several stages.

Let us suppose that we are considering the structure of causal relations in a particular system of kind K in a certain interval of time $[t_o—t_n]$. The following definition captures the way in which a causal model determines the fundamental respects of similarity that are relevant to determining whether one condition makes a difference to another.

> (12) *A model <K, L> of an actual system of kind K generates a sphere of normal worlds* that consists of all and only worlds w such that:
> (i) w contains a counterpart to the actual system and this counterpart has exactly the same K-determining intrinsic properties and relations as the actual system at time t_o;

(ii) $\underline{w}$ does not contain any interfering factors (with respect to the kind $\underline{K}$ and laws $\underline{L}$) during the interval $[t_o—t_n]$;
(iii) $\underline{w}$ evolves in accordance with the laws $\underline{L}$ during the interval $[t_o—t_n]$.

For each sphere of normal worlds, there is a conjunctive proposition that is true of all and only the worlds in the sphere. I shall label this conjunctive proposition $\underline{F}_{M,}$ and, taking over terminology reintroduced by Mackie (1974: p.35) say that it specifies a *field of normal conditions* (generated by the model $\underline{M}$).

Each of the worlds in the sphere of normal worlds generated by the model $\underline{M}$ exemplifies a course of evolution that is *normal*, in a certain sense, for a system of the kind $\underline{K}$ evolving in accordance with the laws $\underline{L}$.[13] The conditions imposed on the these worlds represent *default* settings of the various variables—the initial conditions, the laws, and the absence of interferers—that can influence the way the system evolves through time. If we are investigating the causal relations in a system of a certain kind, as it evolves through a given interval of time, it is reasonable to assume that the initial conditions of the system are the kind-determining intrinsic properties and relations that the system possesses at the beginning of the interval; that the system evolves in accordance with the laws governing the kind of system in question; and that none of the factors that can interfere with the lawful evolution of the system are present. These are default settings in the sense that the assumption that they obtain constitutes a reasonable starting point for our causal investigations, an assumption that we relinquish only when forced to do so. This is not to say that we are always aware of what these default settings are. As mentioned above, there is no reason to think that we will always have complete knowledge of all the initial conditions of a given system, or all the laws governing systems of that kind, or all the possible interferers that can hinder the given system's lawful evolution. Nonetheless, we move from the assumption that the actual system will evolve in accordance with these default settings, whatever their precise details, only when we have good reason to think it must deviate from them.

The normal worlds generated by a model are those that form the background to any consideration of whether some factor makes a difference to another. These normal worlds may hold fixed, as part of the field of conditions, intrinsic properties and relations of the system that are causally relevant to the effects displayed by the system. A special case is that in which the system does not contain any such causally relevant factors. In Newtonian mechanics, a system subject to no forces at all is such a special case. The zero-force law of Newtonian mechanics—the first law of motion—tells us that such a system will remain at rest or travel at a constant velocity. Similarly, a special case in population genetics is a population subject to no evolutionary forces. The evolution of gene frequencies in such a population is described by the zero-force Hardy-Weinberg law. In contrast to these special cases, the typical case is one in which the initial conditions of the system already entail that the system is subject to certain forces. The very description of a Newtonian system consisting of two particles with certain masses and a certain distance apart will

entail that it is subject to gravitational forces. And the very description of a population whose members have certain properties entailing differential fitnesses will ensure that the population is subject to the force of natural selection.[14] Even when a system already possesses an array of causal forces, it makes sense to ask about the causal significance of additional causal forces. The condition for difference-making provides us with a test of the causal significance of these extra factors.

The sphere of normal 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 actual laws. Nonetheless, it is important to note that the actual world need not itself belong to the sphere of normal worlds. For these worlds represent how the actual system would evolve in conformity with the laws *in the absence of any interferers* . In many cases, therefore, these worlds are ideal ones. The actual world, as we know, may be very far from ideal in that the evolution of the actual system may be subject to many interfering forces. The presence of any interfering factor disqualifies the actual world from belonging to the sphere of normal worlds. It follows from this that the Centering Principle that Lewis (1973: pp.26-31) imposes on the similarity relation for counterfactuals fails to hold here, both in its strong and its weak forms. Its strong form states that there is no world as similar to the actual world as the actual world itself, so disallowing ties for most similar world. Its weak form, which allows for ties for most similar world, states that there is no world more similar to the actual world than the actual world itself. The fact that the actual world need not belong to the sphere of normal worlds generated by a model means we must abandon the Centering Principle in both its forms.

So far, we have attended to the question of which worlds count as the normal worlds generated by a model. But definition (7) of a difference-making factor $\underline{C}$ requires a specification of the most similar $\underline{C}$-worlds and the most similar $\sim\underline{C}$-worlds. It is best to spell out the definition of the most-similar $\underline{C}$-worlds by considering sub-cases.

One sub-case we need not consider is that in which both the conditions $\underline{C}$ and $\sim\underline{C}$ are consistent with the field of conditions $\underline{F}_M$. This case cannot arise because it is self-contradictory. For both $\underline{C}$ and $\sim\underline{C}$ to hold consistently with the field of conditions, $\underline{C}$ would have to hold in some of the normal worlds and $\sim\underline{C}$ would have to hold in other normal worlds. But since the laws governing these worlds are deterministic, these worlds would have to differ either with respect to their initial conditions, or with respect to the presence of interfering factors. In either case, the worlds could not satisfy all the conditions (12(i)-(iii)) required for membership in the sphere of normal worlds.

The first sub-case we need to consider—I shall call it Sub-case I—is the case in which the field of conditions $\underline{F}_M$ implies the putative difference-making factor $\underline{C}$. In this kind of case, the initial conditions of the system and the laws, in the absence of interfering factors, imply that the factor $\underline{C}$ holds in the system. As an illustration, consider a slight modification of Mackie's example: a

specially designed machine swings a hammer so that it strikes a chestnut directly from above. Suppose we are considering whether the hammer's striking the chestnut (C) makes a difference to the flattening of the chestnut (E), where the hammer strike is an outcome of the lawful evolution of the relevant system from its initial conditions.

In this kind of case, it is simple to specify which worlds are to count as the most similar C-worlds. They are simply the C-worlds that belong to the sphere of normal worlds; ie those C-worlds that hold fixed the field of conditions F_M. A complication arises, however, when it comes to specifying which worlds are to count as the most similar ~C-worlds. Clearly, ~C is not consistent with F_M, and so the normal worlds included in F_M are not eligible to be the most similar ~C-worlds.

In order to work out which are the most similar ~C-worlds in these circumstances, we need to specify the worlds that differ from the normal worlds no more than is necessary to allow for the realisation of ~C. In other words, we must find the minimal revision of the field of conditions F_M that is consistent with ~C. There are three different elements that determine F_M: the initial conditions of the system, the laws governing the system, and the absence of interfering factors. We can get a set of most similar ~C-worlds by systematically revising each of these elements to allow for the realisation of the counterfactual ~C. And each of the resulting revisions counts, in some sense, as a minimal ~C-inducing revision of F_M. Indeed, each of these revisions generates a similarity relation that corresponds to a certain style of counterfactual reasoning.

For example, we could revise the field of conditions F_M to allow for ~C by revising the laws that govern the system in question while holding fixed the initial conditions and the absence of interferers. Evidently, this kind of revision is required to entertain counterlegals such as 'If force were given by mass times velocity, then...'. However, this kind of revision is not relevant in the present context, in which we are treating counterfactual antecedents that concern particular matters of fact. Another possibility, more relevant in the present context, is to revise F_M by altering the initial conditions while holding fixed the laws and the absence of interferers. This type of revision corresponds to the style of counterfactual reasoning by which we infer how the past conditions must have been different in order for some counterfactual antecedent to be true. This kind of backtracking reasoning lies behind a counterfactual such as 'If the hammer had not struck the chestnut, then the operating machine would have had a malfunction of some kind'. However, the one thing we know from counterfactual analyses of causation is that the required similarity relation must not allow for backtracking reasoning of this kind, on pain of generating countless instances of spurious causation.[15]

The only option left open is to revise F_M by allowing for the presence of an interfering factor that would realise the counterfactual antecedent ~C, while holding fixed the initial conditions and laws of the system. In other words, the most similar ~C-worlds are like the worlds stipulated in (12) above, in that they preserve the initial conditions of the actual system (condition (i)) and the laws

governing the system (condition (iii)), but differ from these worlds in that they allow for an interfering factor that realises the counterfactual antecedent ~$\underline{C}$ (*not* condition (ii)). As discussed above, the paradigm of such an interfering factor is an external human intervention in a system. In small-scale systems open to human manipulation, the kind of interference that would realise a counterfactual antecedent is to be understood in terms of a human intervention. For example, in the modified Mackie example, the most similar worlds that make it true that the hammer does not strike the chestnut are easily imagined: they are simply worlds in which the relevant machine runs on from its initial conditions in conformity with the relevant laws, but at some point a human agent intervenes to prevent the hammer from striking the chestnut. With large-scale systems not open to human manipulation, the interference that realises the counterfactual antecedent can be understood in terms of a miracle that interrupts the lawful evolution of the system.[16] But even here, I would argue, the analogy with human intervention guides the way we think in these cases about the miraculous realisation of the counterfactual antecedents.

It is useful at this point to be able to specify which worlds are to count as the most similar $\underline{C}$-worlds for any antecedent $\underline{C}$, whether or not it is entailed by the field of conditions $\underline{F}_M$. In order to be able to do this, we need an ordering of spheres of worlds in terms of their similarity to the normal worlds in $\underline{F}_M$. (Compare Lewis 1973: pp. 13-16.) If the ordering is to carry information about similarity to the normal worlds, it must satisfy certain conditions.

> (13) Let $\{S_0,..., S_n\}$ be an ordered set of spheres of worlds. This set is *centred* on the sphere of normal worlds S_0 ($=\underline{F}_M$) if and only S_0 is included in every other sphere. The set is *nested* if and only for any spheres S_i and S_j in the set, either S_i is included in S_j or S_j is included in S_i.

When the ordered set of spheres is centred and nested in this sense, it can convey information about the similarity of worlds to the normal worlds. A particular sphere around the sphere of normal worlds will contain just those worlds that resemble the normal worlds to a certain degree. The different spheres will correspond to different degrees of similarity to the normal worlds. The smaller the sphere, the more similar to the normal worlds will be a world falling within it. In other words, if one world falls within a sphere and another world lies outside that sphere, the first world will resemble the normal worlds more closely than the second.

This purely formal specification of the ordering of spheres answers some questions of logic. However, if it is to be applied to particular examples, it must be made more specific with a detailed description of the respects of similarity to the normal worlds that receive significant weighting in the interpretation of conditionals. A complete description of these weightings would require an extensive discussion. However, it will suffice for our treatment of the particular examples of this paper to note one important principle that seems to govern our intuitive judgements about this matter.

(14) *Weightings of Similarity Principle*: In determining the respects of similarity to the normal worlds generated by a model M, it is of first importance to preserve the initial conditions and the laws of the relevant kind of system; and it is of second importance to preserve the absence of interfering factors.

One obvious implication of this principle is that it allows us to read certain counterfactuals in the characteristic non-backtracking manner. It permits a counterfactual antecedent to be realised in a world by an external intervention in the relevant system if the laws and initial conditions of the system are preserved in that world. The principle has another implication that will be relevant to our discussion. We obviously entertain counterfactuals whose antecedents concern changes in the initial conditions of a system. It is perfectly intelligible to say about the modified Mackie example, for instance, 'If the initial conditions of the hammer-striking machine had been different, then the situation would have evolved differently'. But the principle at hand tell us that we have to go further out from the normal worlds to find worlds that permit this counterfactual antecedent than we have to go to find worlds that permit the counterfactual antecedent of 'If the machine's hammer had not struck the chestnut, it would not have been flattened'. Both antecedents can be realised in a world that permits an external intervention in the system. But a world that realises the first antecedent will, of necessity, involve a change in the initial conditions of the system, whereas a world that realises the second antecedent will not. The Weightings Principle implies that the first world must be less similar to the normal worlds than the second world.

With this ordering in hand, we can define the most similar C-worlds in a perfectly general way that covers the case in which C entailed by the field of conditions F_M and the case in which it is not entailed.

(15) The *most similar C-worlds* generated by a model M are the C-worlds that belong to the smallest C-permitting sphere in the ordering of spheres governed by the Weightings Principle.

This is perfectly general also in that it covers, not just Sub-case I in which C is implied by the field of conditions F_M, but also the yet-to-be-considered Sub-case II, in which it is ~C rather than C that is implied by F_M. In this second sub-case, the most similar ~C-worlds are simply the ~C-worlds belonging to the sphere of normal worlds. However, to find the most similar C-worlds in this sub-case, we have to go out from the sphere of normal worlds to find worlds that allow for the realisation of C by intervention or miracle.

We are finally in a position to explicate the idea of one factor making a difference to another in a way that acknowledges the relativity to models.

(16) C *makes a difference to* E in an actual situation relative to the model M of the situation if and only if every most similar C-world generated by the

model is an E-world and every most similar ~C-world generated by the model is a ~E-world.

Of course, these truth conditions bear an unsurprising resemblance to the standard truth conditions for counterfactuals. If the truth conditions for counterfactuals are relativised in a way to match those given above, then the condition can be reformulated to yield the following one:

(17) C *makes a difference to* E in an actual situation relative to the model M if and only if $C \mathrel{\Box\!\!\rightarrow}_M E$ and $\sim C \mathrel{\Box\!\!\rightarrow}_M \sim E$.

Here the subscript M on the counterfactual operator signifies that the operator is defined with respect to the ordering of spheres generated by the model M.

This counterfactual construction is very similar to the notion of counterfactual dependence that plays the central role in Lewis' counterfactual analysis. Indeed, it will be useful to be able to take over this terminology. But the way I will use the term is different from the way Lewis uses it in two respects. First, the counterfactuals that define counterfactual dependence do not, for the reasons given above, conform to the Centering Principle that Lewis imposes on counterfactuals. On the other hand, I believe that they conform to the Limit Assumption to the effect that there is a smallest sphere of antecedent-permitting worlds for any entertainable antecedent. Lewis considers this an optional principle for counterfactuals. But, in fact, it applies automatically to the counterfactuals I have defined, since it follows from the way in which a model generates an ordering of spheres of worlds centred on the normal worlds.

The other way in which the present definition of counterfactual dependence differs from Lewis' is the obvious one bearing on the relativity to a model. The notion of counterfactual dependence, as I will use it, inherits the relativity to a model of the counterfactuals that define it. The truth conditions of counterfactuals in my theory are defined over the most similar antecedent-worlds generated by a model. Lewis' notion involves no such relativity. For he assumes that there is only one kind of system to consider—the whole universe—and so his worlds are maximal worlds. He also assumes that worlds are governed by exceptionless laws without *ceteris paribus* conditions, and so he makes no use of the notion of an interferer in a system which I believe is required to explain the content of *ceteris paribus* conditions. There are further differences between the accounts, but they follow from these.

Let me conclude this section by connecting up our recent discussion with the earlier discussion of the idea that motivates analyses of causation in terms of making a difference. As we have seen, Mackie argues for a conceptual analysis of a cause as 'what makes the difference in relation to some assumed background or causal field' . This idea is best understood, he argues, in terms of the experimental-and-control contrast, rather than the before-and-after contrast. The former contrast can be captured by a pair of conditionals, with

one conditional corresponding to the experimental case and the other conditional corresponding to the control case.

We can now see how to make sense of Mackie's claims. Let us suppose that a conditional is an 'experimental' conditional if we do not have to leave the sphere of normal worlds to find the most similar antecedent-worlds. In other words, the antecedent would be realised by allowing the system in question to evolve lawfully without interference. On the other hand, let us suppose that a conditional is a 'control' conditional if we do have to leave the sphere of normal worlds to find the most similar antecedent-worlds. Or in other words, the antecedent would be realised only by an intervention in the lawful evolution of the system in question from its initial conditions. Given this terminology, we can see that one of the conditionals that define the difference-making condition (17) will be an 'experimental' conditional and the other a 'control' conditional.

It is important to realise, though, that the two conditionals in (17) do not always line up in the same way with the experimental and control cases. It depends on whether we are considering Sub-case I or Sub-case II. In Sub-case I, the 'experimental' conditional is $\underline{C} \mathrel{\Box\!\!\rightarrow}_M \underline{E}$ and the 'control' conditional is $\sim\underline{C} \mathrel{\Box\!\!\rightarrow}_M \sim\underline{E}$. An example in which both these conditionals are true is represented in Figure 1 below. In this figure, the concentric circles represent the spheres of worlds, with the smallest sphere $\underline{F}_M$ representing the sphere of normal worlds generated by a model $\underline{M}$. The symbol @ denotes the actual world.

[Insert Figure 1]

In Sub-case II, $\underline{C} \mathrel{\Box\!\!\rightarrow}_M \underline{E}$ is the 'control' conditional and $\sim\underline{C} \mathrel{\Box\!\!\rightarrow}_M \sim\underline{E}$ the 'experimental' conditional. An example in which both these conditionals are true is represented in Figure 2. (Notice that in both sub-cases $\underline{C} \mathrel{\Box\!\!\rightarrow}_M \underline{E}$ is a factual conditional in the sense that $\underline{C}$ and $\underline{E}$ both actually hold, while $\sim\underline{C} \mathrel{\Box\!\!\rightarrow}_M \sim\underline{E}$ is a genuine counterfactual conditional since $\sim\underline{C}$ and $\sim\underline{E}$ actually fail to hold. Hence, the experimental/control dichotomy crosscuts the factual/counterfactual dichotomy.)

[Insert Figure 2]

Finally, I wish to make a connection with another classic discussion of causation. In their work on causation in the law, Hart and Honoré claim that the concept of a cause as making a difference has its home in a certain paradigm situation. They write:

> 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 kind of 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 which 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 would normally take place, is central to the commonsense concept of cause, and at least as essential as the notions of invariable or constant sequence so much stressed by Mill and Hume. 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 the outcome. (Hart and Honoré 1985: p.29)

Again it is possible to see the point of Hart and Honoré's remarks in the light of the framework developed above. They are, in effect, considering, as a paradigm case, the kind of causal situation that falls under Sub-case II. The normal course of events for a system of some kind, free from any external interference, makes ~E true. If it turns out that E actually holds, then an explanation is required in terms of some factor that makes a difference to this normal course of events. It will be some factor C such that both the 'experimental' conditional $\sim C \mathrel{\Box\!\!\rightarrow_M} \sim E$ and the 'control' conditional $C \mathrel{\Box\!\!\rightarrow_M} E$ are true. By the definition of these conditionals, C will count as an interfering factor in the normal course of development of the system, the kind of interfering factor that is paradigmatically exemplified by an external intervention in the system by an agent. My only reservation about the quotation from Hart and Honoré is that it focuses attention on just one of the two important sub-cases of the idea of making a difference, ignoring the important Sub-case I.

7. The Phenomena Explained

Let us return to the various puzzle cases that were cited as problems for Lewis' account of difference-making in section 3. As we saw in that section, Lewis' account blurs the commonsense distinction between causes and conditions. Let us see whether the present account of difference-making can explain this distinction.

Most philosophical discussions of the distinction between causes and conditions take it for granted that the distinction is best elucidated in a specific explanatory setting. (See, for example, Hart and Honoré 1985: pp.32-44; and Mackie 1974: pp.34-7.) The setting is one in which some unexpected factor E stands in need of explanation, specifically in terms of something that differentiates it from the normal situation in which ~E obtains. This assumption makes sense in terms of the present framework. It simply amounts to the supposition that the causes/conditions distinction is best understood in terms of examples falling under Sub-case II. In examples of this kind, the field of normal conditions F_M generated by a model entails ~E, so that when E

unexpectedly obtains, it requires explanation in terms of a difference-making factor. Let us proceed on the assumption of this explanatory setting, though always keeping in mind that this is just one of two possible sub-cases.

It is interesting to note in this connection a certain abstract implication of the logic of difference-making. It follows from the description of the assumed explanatory setting as exemplifying Sub-case II that the factor identified as making the difference to E must be identified as an interfering factor in the system. The logic of the situation implies that the factor C that makes the difference to E cannot intersect with the field of normal conditions F_M, so that the smallest C-permitting sphere of worlds includes worlds in which C is realised by way of an external interference or intervention in the system. (Figure 2 represents the situation of the explanatory setting.) This implication explains the way in which commonsense and scientific explanations of abnormal occurrences in systems often describe the difference-making factor as an interferer or intervention in the system. Such factors are seen as intrusions into the system that accounts for the deviation from the normal course of events.[17]

In terms of the explanatory setting we are assuming, it is easy to identify the conditions for a causal judgement made relative to a model M. They are simply the conditions that belong to the field of normal conditions F_M. Typically speaking, they will be conditions relating to the initial state of the system in question, conditions relating to the absence of interferers, and any conditions that follow from these and the laws governing the system. Notice that these conditions are not restricted to ones obtaining contemporaneously with the difference-making factor. To modify an example from Hart and Honoré (1985: p.39): if lightning starts a fire in the grass, and shortly after, a normal gentle breeze gets up and the fire spreads to a forest, then the lightning caused the forest fire and the breeze was a mere condition of it. Even though the breeze was subsequent to the lightning, it can be identified as a condition so long as it arises lawfully from the initial conditions of the relevant system and is not an interference in the system.

It follows from this identification of the conditions that no condition can be cited as a difference-maker for the unexpected or abnormal event E in the explanatory setting under consideration. A difference-maker C must be an actually obtaining factor such that all the most-similar C-worlds are E-worlds and all the most-similar ~C-worlds are ~E-worlds. Figure 2, which illustrates this explanatory setting, shows that no actual condition that is entailed by F_M can meet this requirement. In particular, any actual condition X entailed by F_M will be such that all the most similar X-worlds are ~E worlds, contrary to what is required. Consequently, the identification of a condition as an actual factor entailed by F_M implies that conditions cannot be difference-making causes. This is exactly as it should be.

Let us make the discussion concrete by reconsidering the cases cited in section 3, starting with the Lung Cancer case (Example 1). The essential step in treating this example is the identification of the causal model that guides our intuitions about it. It is natural to specify the relevant model here as involving a

person living according to the laws of normal healthy functioning. If the person gets lung cancer, then it is reasonable to look for some factor that makes the difference to this effect with respect to the normal worlds generated by this model—some factor such as his smoking, for instance. Yet, even when the model is specified in these broad terms, we can see that the field of normal conditions generated by this model will hold fixed, as initial conditions, the fact that the person has been born and has lungs. Hence, these conditions cannot be cited as a difference-making factors.

The situation is represented diagrammatically below in Figure 3. The field of normal conditions $\underline{F}_M$ generated by the relevant model $\underline{M}$ entails that the person in question does not get lung cancer ($\underline{LC}$). However, as things actually turns out, the person develops lung cancer, so that an explanation is required in terms of a difference-maker. The figure shows that person's smoking ($\underline{S}$) is such a difference-maker since both of the conditionals $\underline{S} \mathrel{\Box\!\!\rightarrow}_M \underline{LC}$ and $\sim\underline{S} \mathrel{\Box\!\!\rightarrow}_M \sim\underline{LC}$ are true. However, the figure also shows that the person's birth ($\underline{B}$) and his possession of lungs ($\underline{L}$) are mere conditions in this example, as they are entailed by the field of normal conditions. As such, they cannot qualify as difference-makers for the person's lung cancer. (In particular, it turns out that the 'wrong' conditionals $\underline{B} \mathrel{\Box\!\!\rightarrow}_M \sim\underline{LC}$ and $\underline{L} \mathrel{\Box\!\!\rightarrow}_M \sim\underline{LC}$ hold, so that $\underline{B}$ and $\underline{L}$ cannot be make the appropriate counterfactual difference to $\underline{LC}$.) The figure also shows that there are outer spheres of worlds, quite dissimilar to the normal worlds of $\underline{F}_M$, that permit the absence of these conditions. In these outer spheres, it is true that $\sim\underline{B} \mathrel{\Box\!\!\rightarrow}_M \sim\underline{LC}$ and $\sim\underline{L} \mathrel{\Box\!\!\rightarrow}_M \sim\underline{LC}$, as one would expect to be the case.

[Insert Figure 3]

Of course, the causal claims about the person can be interpreted in terms of a different causal model. By changing the example, we can make a different causal model more salient. Consider an example given by Lewis in which we are to suppose that there are gods who take a keen interest in human affairs.

> It has been foretold that the event of your death, if it occurs, will somehow have a momentous impact on the heavenly balance of power. It will advance the cause of Hermes, it will be a catastrophe for Apollo. Therefore Apollo orders one of his underlings, well ahead of time, to see to it that this disastrous event never occurs. The underling isn't sure that just changing the time and manner of your death would suffice to avert the catastrophe; and so decides to prevent your death altogether by preventing your birth. But the underling bungles the job: you are born, you die, and it's just as catastrophic for Apollo as had been foretold. When the hapless underling is had up for charges of negligence, surely it would be entirely appropriate for Apollo to complain that your birth caused your death. (Lewis 1999: p.35)

It would, indeed, be appropriate for Apollo to make this causal claim. But notice how the causal model has changed, with a different kind of system and different laws of functioning involved. Indeed, this example is better seen as exemplifying Sub-case I rather than II, as the field of normal conditions $\underline{F}_M$ generated by the relevant model entail the factor to be explained, namely that you will die (D). This is represented diagrammatically below in Figure 4. Nonetheless, your birth (B) does make a difference to your death in view of the fact that Apollo's underling could have intervened in the system (I) to prevent your birth and so your death. Here the two counterfactuals required for your birth to be a difference-maker for your death, that is $\underline{B} \mathrel{\Box\!\!\rightarrow}_M \underline{D}$ and $\sim\underline{B} \mathrel{\Box\!\!\rightarrow}_M \sim\underline{D}$, both hold true with respect to the sphere of normal worlds generated by the relevant model. But it does require a radical change of causal model to get this result.

[Insert Figure 4]

The present framework provides a ready explanation of the two forms of contextual relativity underlying the commonsense causes/conditions distinction. It is easy to see how the explanation should work for the Presence of Oxygen case (Example 2). Here our readiness to rank the presence of oxygen as a condition in one situation and a cause in another simply reflects the fact that the two situations involve different kinds of systems with different initial conditions. In the first situation, in which a fire takes hold of a building, the presence of oxygen is an initial condition, held fixed in the field of normal conditions; whereas in the second situation, in which oxygen is excluded from a delicate experimental or manufactory set-up, it is not an initial condition, so it is eligible to be a difference-maker for the effect.

The explanation of the relativity of causes/conditions distinction to the context of enquiry is equally straightforward. In the Ulcerated Stomach case (Example 3), the causal claims made by the different enquirers are explained by the fact that they are employing different models. For example, the person who prepares meals for the patient implicitly employs a model that focuses on the person with the ulcerated stomach as a fixed initial condition. On the other hand, the doctor tacitly employs a model that focuses on the person as a normally functioning human without ulcerated stomach as a fixed initial condition. These enquirers both seek factors that make a difference to the patient's indigestion, but they do so with respect to the different spheres of normal worlds generated by their different models.

Another virtue of the present framework is that it allows us to discriminate between absences as causes and absences as mere conditions. In allowing absences, and other non-occurrences as causes, we need not open the floodgate to a host of spurious causes. For instance, the natural causal model for interpreting the Multiple Omissions case (Example 5) ranks the gardener's omission as the cause of the plant's death, while backgrounding other people's omissions as mere conditions. This causal model is one that takes, as its system for investigation, a healthy plant functioning under a regime of regular

watering by the gardener. Accordingly, this example can be seen to exemplify Sub-case II, in which the field of normal conditions $\underline{F}_M$ entails that the gardener waters the plant (GW) and the plant survives (PS). The situation is represented in Figure 5 below. When the plant fails to survive, an explanation in terms of a difference-maker is required. The gardener's actual omission can act as such a difference-maker, since the two appropriate counterfactuals hold true, $\underline{GW} \mathrel{\square\!\!\rightarrow_M} \underline{PS}$ and $\sim\underline{GW} \mathrel{\square\!\!\rightarrow_M} \sim\underline{PS}$. (Notice that in this case what explains the abnormal occurrence of the plant's death is actually a non-occurrence. It seems that commonsense sometimes regards a non-occurrence as an interfering factor that perturbs the normal course of development of some kinds of systems.) However, the omission by everyone else ($\sim\underline{SW}$) to water the plant is disqualified from acting as a difference-maker, as this omission is held fixed in the field of normal conditions. (However, an outer sphere allows it to be the case that someone else waters the plant and in this outer sphere it holds true that $\underline{SW} \mathrel{\square\!\!\rightarrow_M} \underline{PS}$.)

[Insert Figure 5]

A similar explanation can be given of our causal judgements about the Absence of Nerve Gas case (Example 4), when it is interpreted as exemplifying Sub-case II. As a matter of fact, this is a rather strained interpretation, as one has to imagine a field of normal conditions generated by an appropriate model that entails that I am not writing at my computer. However, if we do imagine this, then given the fact that I am so writing, it is reasonable to ask for some explanation in terms of a difference-maker. But this difference-maker cannot be supplied by the absence of nerve gas, the absence of flamethrower attack, or absence of meteor strike. For these absences are held fixed in the field of normal conditions, in view of the fact that the intrusion of nerve gas, or flamethrower attack, or meteor strike would count as an interference in the system.

It is much easier to understand this example as exemplifying Sub-case I; that is, the field of normal conditions generated by the relevant model entails the actually obtaining factor—my writing at my computer. However, a striking fact emerges when one construes the example in this way. The various absences mentioned above can each count as a difference-maker for my writing. For instance, the intrusion of nerve gas into the situation in which I am writing at my computer is naturally regarded as an interference, whose absence should be held fixed in the field of normal conditions. However, the rules for a model's generating spheres of worlds (in particular, the Weightings Principle) permit the presence of the nerve gas in an outer sphere of worlds. The consequence of this is that the two counterfactuals required for the absence of nerve gas ($\sim\underline{NG}$) to count as a difference-maker for my writing ($\underline{W}$) can both hold. Figure 6 below represents the situation in which the two counterfactuals $\underline{NG} \mathrel{\square\!\!\rightarrow_M} \sim\underline{W}$ and $\sim\underline{NG} \mathrel{\square\!\!\rightarrow_M} \underline{W}$ hold. The same line of reasoning shows that the absence of any factor that could be regarded as an interferer can, in the right circumstances, act as a difference-maker for my writing at my computer, when

the example is construed as exemplifying Sub-case I rather than II. I conjecture that it is not absurd to judge in these circumstances that a *sustaining cause* of my writing at my computer is the collective absence of all interfering factors, including the absence of nerve gas, the absence of flamethrower attack, and the absence of meteor strike.

[Insert Figure 6]

In the discussion above, I have argued that the commonsense distinction between causes and conditions makes sense only relative to a field of normal conditions generated by a causal model. Given such a model, the conditions of some effect can be explained as those factors belonging to the field, and the causes as those factors that make the difference to the effect relative to this field. It is worth comparing this characterisation of the distinction with an alternative one that has become popular. On this alternative characterisation, the distinction is an entirely pragmatic one to be cashed out in terms of contrastive explanation. We have already seen the outlines of this kind of approach in section 4. The main idea is that commonplace causal judgements are implicit answers to contrastive why-questions of the form 'Why does $\underline{E}_1$, rather than $\underline{E}_2$, …, $\underline{E}_n$, obtain?', where the members of the contrast class {$\underline{E}_1$, …, $\underline{E}_n$} may or may not be mutually compatible. On this approach, a cause is an actual 'objective cause' that differentiates $\underline{E}_1$ from the other members of the contrast class; and the conditions are those factors that are common to all possible situations that could realise a member of the contrast class. This approach also makes the distinction a context-sensitive one because different contexts may contrast $\underline{E}1$ with different sets of alternatives, so affecting which factors count as causes and conditions.

There are, to be sure, similarities between these characterisations of the cause/conditions distinction. One obvious similarity is that they both characterise the distinction in terms of a contextually generated space of possibilities. In the present framework, it is the space of normal worlds generated by a model; in the alternative framework, it is to the space of contrasting alternatives. Still, there are, in my view, some serious shortcomings to this alternative characterisation, some of which have been touched on earlier.

One of these is that the characterisation must operate with an independently motivated notion of an 'objective cause'. For the factor that differentiates $\underline{E}_1$ from the other members of the contrast class cannot be a causally irrelevant factor: it must be an 'objective cause' present in the $\underline{E}_1$ situation, but not in the others. But this requires an explication of what an 'objective cause' is. The explication cannot, on pain of unnecessary duplication, appeal to the idea of a difference-maker. Another shortcoming of the characterisation is that it leaves radically under-specified what conditions are common to the realisations of the different members of the contrast class. A specification of these commonalities requires a description of a similarity relation between the possible situations realising the different members of the contrast class. It is totally unclear what this similarity relation involves. Lewis

attempts, as we have seen, to specify a similarity relation by appeal to backtracking reasoning, but this attempt fails to deliver determinate verdicts in many cases.

For these reasons, I believe, the alternative characterisation of the causes / conditions distinction in terms of contrastive explanation cannot bear the explanatory weight that many have placed on them. Why then, it may be asked, does the account of the phenomena in terms of contrastive explanation, sketched in section 4, work as well as it does? There are several reasons, I would suggest. First, the treatment of causal judgements as answers to contrastive why-questions puts the emphasis in the right place, namely on the context-relativity of these judgements. Secondly, the specification of the contrast class, embodied in a contrastive why-question, carries information about the kind of system that is being investigated and its laws of normal functioning. In other words, we can read off from a class of contrasting alternatives information about the real contextual determinant of our causal judgements—the underlying causal model. But the contrast class is, at best, an indirect source of this information.

So, I oppose the popular strategy of explaining the commonsense view of causes as difference-makers pragmatically in terms of contrastive explanation. Rather, I recommend the reverse procedure of explaining contrastive explanation in terms of the present independently motivated account of difference-making. Let me outline in broad detail how such an explanation should work. Suppose the contrastive why-question 'Why does $\underline{E}_1$, rather than $\underline{E}_2$, …, $\underline{E}_n$, obtain?' has been posed, where the members of this contrast class are all actual or possible factors of systems of kind $\underline{K}$ operating according to laws $\underline{L}$. Then, a satisfactory answer to this question should cite some actual factor $\underline{C}$ that makes a difference to $\underline{E}_1$ relative to the model $\underline{M} = \langle \underline{K}, \underline{L} \rangle$. Where the members of the contrast class are *incompatible* outcomes in the same system, it follows from the definition of $\underline{C}$ as a difference-maker for $\underline{E}_1$ that we have an automatic explanation why none of the alternative possible outcomes could have occurred. Similarly, where the members of the contrast class are *compatible* outcomes of different systems of kind $\underline{K}$, it follows from the definition of $\underline{C}$ as a difference-maker for $\underline{E}_1$ that we have an automatic explanation why none of the alternative outcomes actually occurred. (For, given that these have the same initial conditions and conform to the same laws, these systems would have to have $\underline{E}_1$ if they had the factor $\underline{C}$.)

This account of contrastive explanation overcomes the difficulties facing Lewis's account that we encountered in section 4. For example, it does not leave it indeterminate what the various members of the contrast class have in common. It specifies these commonalities precisely in terms of the field of normal conditions generated by the relevant causal model. Again, it handles the examples such as Hempel's, where a contrastive explanation is required of two compatible alternatives. We explain why Jones, rather than Smith, got paresis, by citing the difference-making factor of syphilis that actually applies to Jones, but not to Smith. Most importantly, this account does not involve an unnecessary duplication of the idea of a difference-maker. It follows from the

definition of difference-making that, if we have factor C that makes a difference to E_1, then we have a contrastive explanation of why E_1 rather than E_2, …, E_n. This factor does its work of differentiating the contrasting alternatives precisely because it is, by hypothesis, a difference-maker for E_1.

8. Conclusion

One of the aims of this paper has been to explore the conception of a cause as 'what makes the difference in relation to some assumed background or causal field.' (Mackie: 1974, p. 71) I have tried to explicate this conception by giving an account of difference-making in terms of context-sensitive counterfactuals. This account explains the way in which we distinguish causes from background conditions in terms of the way in which an implicit contextual parameter of a causal model generates a similarity ordering among possible worlds. It is important to elucidate this dimension of context-sensitivity in our causal judgements not only to get the conceptual analysis of causation right, but also to avoid philosophical puzzles that arise from too simplistic conceptions of causation. For example, the puzzle in the philosophy of mind about the mental causation—the puzzle of how mental states can play a role in the causation of behaviour independent of the role played by the physical states on which they supervene—arises because philosophers overlook the way in which causal models implicitly guide our judgements about causation, or so I argue (in Menzies 2002).

References

Anderson, J. 1938: 'The Problem of Causality', *Australasian Journal of Philosophy and Psychology*, 41.

Cartwright, N. 1983: *How the Laws of Physics Lie*. Oxford: Clarendon Press.

Cartwright, N. 1999: *The Dappled World*. Oxford: Oxford University Press.

Dretske, F. 1973: 'Contrastive Statements', *The Philosophical Review*, 82, pp. 411-37.

Ducasse, C.J. 1968: *Truth, Knowledge, and Causation*. London: Routledge Kegan Paul.

Fodor, J. 1991: 'You Can Fool Some of the People All the Time, Other Things being Equal: Hedged Laws and Psychological Explanation', *Mind*, 1000, pp. 19-34.

Garfinkel, A. 1981: *Forms of Explanation*. New Haven: Yale University Press.

Giere, R. 1988: *Explaining Science*. Chicago: University of Chicago Press.

Gorovitz, S. 1965: 'Causal Judgements and Causal Explanations', *Journal of Philosophy*, 62, pp. 695-711.

Grice, H. P. 1975: 'Logic and Conversation' in P. Cole and J.L. Morgan (eds.) *Syntax and Semantics*, Volume 3, Academic Press.

Hart, H. L. and Honoré, A. 1985: *Causation in the Law*. 2nd edition. Oxford: Clarendon Press.

Hempel, C. 1965: *Aspects of Scientific Explanation*. New York: The Free Press.

Hitchcock, C. 1996: 'The Role of Contrast in Causal and Explanatory Claims', *Synthese*, 107, pp. 395-419.

Joseph, G. 1980: 'The Many Sciences and the One World', *Journal of Philosophy*, 77, pp. 773-90.

Kim, J. 1982: 'Psychophysical Supervenience', *Philosophical Studies*, 41, pp. 51-70.

Lewis, D. 1973a: *Counterfactuals*. Oxford: Basil Blackwell.

Lewis, D. 1973b: 'Causation', *Journal of Philosophy*, 70, pp. 556-67. Reprinted in Lewis 1986.

Lewis, D. 1979: 'Counterfactual Dependence and Time's Arrow', *Nous*, 13, pp. 455-71. Reprinted in Lewis 1986.

Lewis, D. 1986: *Philosophical Papers, Volume II*. Oxford: Oxford University Press.

Lewis, D. 1999a: 'Causation as Influence', University of Melbourne Preprint 1/99. Reprinted in this volume.

Lewis, D. 1999b: *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.

Langton, R. and Lewis, D. 1999: 'Defining "Intrinsic"', reprinted in Lewis 1999b.

Lipton, P. 1990: 'Contrastive Explanation', in D. Knowles (ed.) *Explanation and its Limits*, Cambridge: Cambridge University Press.

Lipton, P. 1991: *Inference to the Best Explanation*. London: Routledge.

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

Menzies, P. 1989: 'A Unified Theory of Causal Relata', *Australasian Journal of Philosophy*, 67, pp. 59-83.

Menzies, P. 1996: 'Probabilistic Causation and the Pre-emption Problem', *Mind*, 105, pp. 85-117.

Menzies, P. 1999: 'Intrinsic versus Extrinsic Conceptions of Causation' in H. Sankey (ed.), *Causation and the Laws of Nature*, Dordrecht: Kluwer Academic Publishers, pp. 313-29.

Menzies, P. 2002: 'The Causal Efficacy of Mental States', in J. Monnoyer (ed.) *The Structure of the World: The Renewal of Metaphysics in the Australian School*. Paris: Vrin Publishers.

Mill, J. S, 1961: *A System of Logic*. London: Longmans.

Pietroski, P. and Rey, G. 1995: 'When Other Things Aren't Equal: Saving *Ceteris Paribus* Laws from Vacuity', *British Journal for the Philosophy of Science*, 46, pp. 11-121.

Schiffer, S: 1991: '*Ceteris Paribus* Laws', *Mind*, 100, pp. 1-18.

Sober, E. 1980: 'Evolution, Population Thinking, and Essentialism', *Philosophy of Science*, 47, pp. 350-83.

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

Suppe, F. 1979: 'Introduction' to *The Structure of Scientific Theories*. Urbana: University of Illinois Press.

Suppe, F. 1989: *The Semantic Conception of Theories and Scientific Realism*. Urbana: University of Illinois Press.
Toulmin, S. 1961: *Foresight and Understanding: An Inquiry into the Aims of Science*, Indianapolis: Indiana University Press.
Unger, P. 1977: 'The Uniqueness of Causation', *American Philosophical Quarterly*, 14, pp. 177-88.
van Fraassen, B. 1980: *The Scientific Image*. Oxford: Clarendon Press.
van Fraassen, B. 1989: *Laws and Syummetry*. Oxford: Oxford University Press, 1989.
Woodward, J. 1984: 'A Theory of Singular Causal Explanation', *Erkenntnis*, 21, pp. 231-62.

---

[1] Versions of this paper have been read at the July 1996 conference of the Australasian Association of Philosophy in Brisbane, and at seminars at the University of Sydney (April 1997), the California Institute of Technology (April 1997), and the Research School of Social Sciences, ANU (June 97). I wish to thank the audiences at these places for their probing questions. I especially wish to thank John Bacon, Nancy Cartwright, Greg Currie, Chris Daly, Phil Dowe, Brian Garrett, Ian Gold, Karen Green, Alan Hajek, Adrian Heathcote, Frank Jackson, Ellie Mason, Michael McDermott, George Molnar, Philip Pettit, and Jim Woodward.

[2] I have considered some of the problems faced by the counterfactual approach to causation in connection with pre-emption cases in the paper Menzies 1996. There I argue that cases of pre-emption show that a purely counterfactual analysis of causation will not work: at some point we must make an appeal to a concept of causation as an intrinsic relation or process in order to deal with them. (See also Menzies 1999.) I show how to marry counterfactual intuitions about causation with intuitions about intrinsic processes by way of a Ramsey-Carnap-Lewis treatment of causation as a theoretical relation. On this treatment, a counterfactual dependence is a defeasible marker of causation: when the appropriate conditions are satisfied, it picks out the process that counts as the causal relation. Such a treatment of causation as a theoretical relation can obviously be framed around the counterfactual explication of the idea of a cause as a difference-maker recommended in this paper.

[3] Actually, Lewis has presented three theories of causation, with the third being the 'quasi-dependence' theory that he tentatively sketched in Postscript E to the paper 'Causation' (Lewis 1986).

[4] Mackie credits the term 'field of causal conditions' to his teacher John Anderson, who used it to resolve difficulties in Mill's account of causation in the paper Anderson 1938.

[5] I take the terms for the different kinds of context relativity from Gorovitz 1965.

[6] It might be argued that these examples only demonstrate that the construction '<u>c</u> is *the* cause of <u>e</u>' is context-sensitive; and that the counterfactual theory is best understood as an account of the context-insensitive construction '<u>c</u> is *a* cause of <u>e</u>'. Thus it might be argued that even the World Food Authority would admit that the drought was *a* cause of the famine and that the doctor would allow that eating parsnips was *a* cause of the indigestion. While this defence on the basis of common usage seems faintly acceptable with some examples, it fails in other cases. Even with a liberal understanding of the words, it seems stretched to say that a person's birth and possession of lungs were among the causes of his lung cancer. It seems that even the expression 'a cause' displays some degree of context sensitivity. For discussion and elaboration of this point see Unger 1977.

[7] Mill wrote: 'Nothing can better show the absence of any scientific ground for the distinction between the cause of a phenomenon and its conditions, than the capricious way in which we select from among the conditions that which we choose to denominate the cause.' See Mill 1961: p. 214.

[8] The clearest exponent of this strategy of adding on a theory of contrastive explanation to a theory of 'objective causation' is Peter Lipton: see Lipton 1990 and 1991: Chapters. 3-5. Lipton does not, however, endorse Lewis' counterfactual theory as the right theory of 'objective causation'. Others who have emphasised the importance of contrastive explanation for understanding ordinary causal discourse include Gorovitz 1965, Mackie 1974, Dretske 1973, Woodward 1984, and Hitchcock 1996.

[9] It is worth observing that the idea of an integrated account of context-sensitivity is not altogether foreign to Lewis' style of counterfactual theory. In the revamped 1999 version of the theory, context enters the theory in an important but inconspicuous way. The notion of a not-too-distant alteration of the cause introduces an important contextual element into the truth conditions of causal statements. A not-too-distant alteration of the cause is an alteration that is relevantly similar to the cause by the standards determined by the context. The approach that I shall advocate is similar in building the context-sensitivity into the truth conditions, but it will draw on contextually determined standards of similarity for counterfactuals rather than events.

[10] I develop a fuller account of factors in Menzies 1989, though the paper uses the term 'situations' rather than 'factors'.

[11] A problem infects these definitions that is parallel to the problem Lewis pointed out for Kim's definition of the simple concepts. Modifying some concepts Lewis introduced, let us say that a system <u>S</u> is *accompanied* if and only if it coexists with some contingent object wholly distinct from it, and *lonely* if and only if it does not so coexist. The definitions I have presented amount to saying that the extrinsic properties of a system are those implied by the

accompaniment of the system and the intrinsic properties of a system are those compatible with its loneliness. The problem is that loneliness of a system is intuitively an extrinsic property of the system (since it can differ between duplicates of the system), but it counts as an intrinsic property by the definition (since it is compatible with itself). One possible remedy to this problem may be to adapt to our purposes the refinement of Kim's original idea to be found in Langton and Lewis 1999. This refinement is supposed to circumvent the defect Lewis detected in Kim's original idea.

[12] The concepts of relations that are extrinsic or intrinsic to a set of objects can be defined in a similar manner. A relation R is *extrinsic* to a set of objects if and only if necessarily, the relation holds between two members of the set only if some other contingent object wholly distinct from the set exists. Conversely, a relation R is *intrinsic* to the set if and only if possibly, the relation holds between two constituents of the system although no contingent object wholly distinct from the set exists.

[13] In using the terms 'sphere of *normal* worlds' and 'field of *normal* conditions', I am not invoking the ordinary notion of 'normal'. Rather the given definitions stipulate the intended sense in which I use the terms, though I hope this sense bears some relation to the ordinary notion of 'normal'.

[14] For an illuminating discussion of the role of zero-force laws, such as the first law of motion and the Hardy-Weinberg law, in default assumptions about the evolution of systems, see Elliott Sober's discussion (1984: Chap. 1).

[15] See, for example, Lewis' discussion of the perils of allowing backtracking counterfactuals in the counterfactual analysis of causation in his Postscripts to 'Counterfactual Dependence and Time's Arrow' and 'Causation' in his (1986).

[16] For an account of the similarity relation for non-backtracking counterfactuals that appeals to the miraculous realisation of antecedents, see Lewis 1979.

[17] For discussions of episodes in the history of science that highlight the significance of causes as intrusions in the normal course of events, see Toulmin 1961: Chapters 3-4; and Sober 1980.