

# **Causation and Counterfactuals**

edited by John Collins, Ned Hall, and L. A. Paul

A Bradford Book  
The MIT Press  
Cambridge, Massachusetts  
London, England

© 2004 Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

This book was set in Times New Roman on 3B2 by Asco Typesetters, Hong Kong, and was printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Causation and counterfactuals / edited by John Collins, Ned Hall, and L. A. Paul.

p. cm. — (Representation and mind)

“A Bradford book.”

Includes bibliographical references and index.

ISBN 0-262-03317-8 (hc. : alk. paper) — ISBN 0-262-53256-5 (pb. : alk. paper)

1. Causation. 2. Counterfactuals (Logic). I. Collins, John David. II. Hall, Edward J. (Edward Jonathan), 1966– . III. Paul, L. A. (Laurie Ann), 1966– . IV. Series.

BD541.C192 2004

122—dc22

2003060612

10 9 8 7 6 5 4 3 2 1

To the memory of David Lewis



# 10 Void and Object

David Lewis

## 1 The Deadly Void

The void is deadly. If you were cast into a void, it would cause you to die in just a few minutes. It would suck the air from your lungs. It would boil your blood. It would drain the warmth from your body. And it would inflate enclosures in your body until they burst.<sup>1</sup>

What I've said is literally true, yet it may be misleading. When the void sucks away the air, it does not exert an attractive force on the air. It is not like a magnet sucking up iron filings. Rather, the air molecules collide and exert repulsive forces on one another; these forces constitute a pressure that, if unresisted, causes the air to expand and disperse; the void exerts no force to resist the pressure; and that is why the air departs from the lungs.

Likewise, when the void boils the blood, there is no flow of energy from the void into the blood. It isn't like a stove boiling a kettle of water. The blood is already warm enough to boil, if its vapor pressure is unresisted; the void exerts no counter-pressure; and so the boiling goes unprevented.

Likewise, when the void drains your warmth, what happens is that your thermal energy, left to itself, tends to dissipate; and the void provides no influx of energy to replace the departing heat.

And when the void inflates enclosures, again what happens is that the enclosed fluids exert pressure and the void exerts no counterpressure. So nothing prevents the outward pressure from doing damage.

In short, you are kept alive by forces and flows of energy that come from the objects that surround you. If, instead of objects, you were surrounded by a void, these life-sustaining forces and flows would cease. Without them, you would soon die. That is how the void causes death. It is deadly not because it exerts forces and supplies energy, but because it doesn't.

## 2 Void and Vacuum

The deadly effects of a void would be just like those of a commonplace vacuum. Nevertheless, I distinguish the two. The more we learn about the vacuum, the more we find out that it is full of causally active objects: force fields, photons, and "virtual" particles. Spacetime itself, if curved, can serve as a repository of energy. And perhaps that is not the end. The void, on the other hand, is entirely empty. Thus, if there is a vacuum within these four walls, there may be quite a lot of objects between

the walls that are capable of exerting forces and supplying energy. Whereas if there is a void within these walls, then (even though the walls are some distance apart) there is nothing at all between the walls. What?—Not even any spacetime? Not even any flat, causally inert spacetime?—No, not even any spacetime. Nothing at all.

The void is what we *used* to think the vacuum was. It is what the vacuum is according to a relational theory of spacetime, according to which particles are surrounded by nothing at all, and are separated not in virtue of substantial spacetime between them but rather by direct external distance relations from one particle to another.

Whether or not any such relational theory is true, I take it to be, in some good sense, a genuine possibility for the world. Nothing rules it out a priori. It is a “broadly logical” or “conceptual” possibility. But conceptual possibility is governed by a combinatorial principle that says that we can generate new possibilities by patching together (copies of) parts of other possibilities.<sup>2</sup> So if a relationist world is possible, and a world full of substantial spacetime is likewise possible, then by patching together parts of these two worlds, we get a world that consists of substantial spacetime interrupted by occasional voids. The walls and the spacetime within them are distinct existences; ergo it is possible for either one to exist without the other. If the walls exist without the spacetime (and without any other objects between the walls) then there is a void between the walls.

A void is conceptually possible. But probably it is impossible in another sense: It violates the laws of nature. Nothing you or nature can do will make a void, but only a vacuum. Our horrific counterfactuals about what would happen if you were cast into a void are contrary not only to fact but to law.

Yet they are none the worse for that. It is no mistake to conflate the deadly void and the deadly vacuum in our counterfactual reasoning, because the difference between the two makes very little difference to the lethal effects. If you were cast into the vacuum of outer space, high-energy photons would be the least of your problems. It is the absences that would do you in. And when it comes to absences, the void is like the vacuum, only more so. Therefore, although presumably we may not entertain the supposition that you are cast into a void under laws that are exactly those of the actual world, we know very well which features of actuality to hold fixed in supposing there to be a void.

### 3 Menzies on Causation as Intrinsic

Peter Menzies has identified a source of dissatisfaction with our most popular approaches to the analysis of causation.<sup>3</sup> We would prefer, he says, to think of the

causal relation as an *intrinsic relation* between cause and effect: a relation that is instantiated by a pair of events just in virtue of the (natural) properties and relations of that pair itself, and so supervenes just on the (natural) properties and relations of the pair; a relation that is independent of the (natural) properties and relations of all things that are entirely distinct from that pair; and hence a relation that would be instantiated equally by any other pair of events that shared exactly those (natural) properties and relations. It would be instantiated equally by any such duplicate pair regardless of whether the duplicate pair was actual or merely possible; regardless of whether the duplicate pair was all alone in its universe or whether it was accompanied by contingent objects distinct from itself; and regardless of what pattern was constituted by the accompanying objects, if there were any.<sup>4</sup>

Two events *c* and *e* stand in a relation of constant conjunction iff, throughout the universe, there are many events of the same kind as *c*, and every one of them is accompanied by an event of the same kind as *e*. An analysis that identifies causation with constant conjunction is subject to many difficulties and admits of many improvements, but one difficulty remains: constant conjunction is not an intrinsic relation, and neither is any of its improved descendants.

Two events *c* and *e* stand in a relation of counterfactual dependence iff, if *c* had not occurred, *e* would not have occurred either.<sup>5</sup> A counterfactual analysis of causation is again in need of improvement. We need to consider a pattern of counterfactuals that goes well beyond the simple counterfactual dependence of *e* on *c*.<sup>6</sup> Whatever this pattern may be, it obtains in virtue of whatever makes single counterfactuals true or false. And in this, a large part is played by the laws of nature. (That is so even when we entertain counterlegal suppositions. We may have to bend the laws, but we do not give them away altogether.) And laws, whatever else they may be, are at least exceptionless regularities throughout the universe (or some large part of the universe). So counterfactual dependence is not an intrinsic relation, and neither is any of its improved descendants.

If we thought that causation ought to be analyzed as an intrinsic relation between events, then constant-conjunction analyses and counterfactual analyses would both be in trouble. The trouble would go deep. It would persist no matter how well we succeeded in the game of counterexamples and improvements.

#### 4 Menzies on Causal Functionalism

What sort of analysis of causation would portray the causal relation as intrinsic? Menzies offers a plausible recipe: he applies the “Canberra plan” to causation.<sup>7</sup> We start with a platitudinous folk theory of the causal relation. This theory says that the

causal relation does so-and-so and such-and-such. Thereby it specifies a functional role that the causal relation is said to occupy. We stipulate that the causal relation is the relation, if there is one, that does in fact occupy the specified role.

(At least, that is the ideal case. But if there is no unique relation that perfectly occupies the role, we might resort to semantic satisficing. We might grant that an imperfect occupant of the role is an imperfect, but good enough, deserver of the name “causal relation,” if it comes near enough and there is no better candidate to be had. Or if several candidates were tied for best, so that they deserved the name equally, we might grant them equal claim by treating the name as ambiguous or indeterminate in reference.)<sup>8</sup>

Menzies lists three “crucial platitudes.” First, the causal relation is a relation between wholly distinct events. Second, it is a relation that is intrinsic to the event-pairs that instantiate it. Third, it is typically, though perhaps not invariably, associated with a probabilistic version of counterfactual dependence. In most cases in which it relates an event  $c$  to an event  $e$ , though perhaps not in a minority of atypical cases, the actual objective single-case probability of event  $e$  is substantially higher than it would have been if  $c$  had not occurred. (For instance, in the most extreme case, the actual probability of  $e$  is exactly 1, without even an infinitesimal chance of not- $e$ ; whereas without  $c$ , the probability of  $e$  would have been exactly 0, without even an infinitesimal chance of  $e$ , so that  $e$  would definitely not have occurred.)

By building the intrinsic character of the causal relation into the definitive functional role as a crucial platitude, Menzies makes sure that the occupant of the role, if there is one (and if it is a perfect occupant) must be an intrinsic relation. Does this strategy satisfy Menzies’s desideratum?

Not entirely. In this case, as in others, the functionalist strategy creates an ambiguity. We have the functional role; we have the relation that is the actual occupant of that role. But presumably it is a contingent matter what relation occupies the role. In a different possible world, perhaps with different laws of nature, some different relation may occupy the role. Now let  $c$  and  $e$  be two events that both occur in some possible world  $w$ . Shall we say that  $c$  causes  $e$  iff  $c$  stands to  $e$  in the relation that occupies the role in actuality? Or iff  $c$  stands to  $e$  in the relation that occupies the role in world  $w$ ? If we make the former decision, then indeed (at least if the actual occupation of the role is perfect) causation is an intrinsic relation. But if we make the latter decision, then causation—by which I mean the relation such that, for any events  $c$  and  $e$  in any world  $w$ ,  $c$  bears it to  $e$  in  $w$  iff  $c$  causes  $e$  in  $w$ —will be a disjunctive affair. It will not be the same as any of the various role-occupants in the various worlds.<sup>9</sup> So even if all these role-occupants are intrinsic relations, it will not follow that causation itself is an intrinsic relation.

The folk well might have left this subtle ambiguity unresolved. Indeed, they might never have noticed it. After all, they are mostly interested in causation as it takes place here in our actual world (or in worlds similar enough to ours that they could be expected to have the same role-occupant). What mostly matters is that the actual occupant of the role is an intrinsic relation. So I think that if Menzies's strategy is otherwise successful, then it satisfies his intuitive desideratum quite well enough. But we cannot say unequivocally that he has *analyzed* causation as an intrinsic relation.

## 5 Causal Functionalism and the Void

Menzies's strategy is not otherwise successful. What purports to be a general analysis of causation turns out to apply only to one kind of causation. Causation by the void, and causation by absences more generally, have no place in Menzies's account. Omission is omitted.

Menzies's topic is the causal relation. A relation requires relata. The void affords no causal relata: There's nothing there at all, so there's nothing for events to happen to, so the void is devoid of events. And even if we allow causal relata to belong to other categories, still there would be none of them in the void—because there's nothing at all in the void. A vacuum, or an almost-void that still contains flat, inert spacetime, is almost as bad. It has parts; and it may contain events in some tolerant sense of that word. But it contains nothing that could plausibly be said to bear a causal relation to the death of a victim. The victim dies, as I said before, not because of what *is* there, but because of what *isn't*. Indeed, whenever any effect is caused by an absence of anything, we have the problem of the missing relatum. (And likewise whenever anything causes an absence.) A void, being the absence of any objects at all, is just the most extreme case of an absence.

Faced with the problem of the missing relatum, we have four possible lines of response.

(1) We could deny, in the face of compelling examples to the contrary, that absences ever cause anything. We could deny, for instance, that the void is deadly. (Likewise, we could deny that anything ever causes an absence. In other words, we could deny that there is any such thing as prevention.) Simply to state this response is to complete the *reductio* against it.

(2) We could reify absences nonreductively. A void, so we might say, is a *sui generis* entity, but it is none the worse for that. It is eligible to serve as a causal relatum. It springs up automatically and necessarily whenever, and only whenever, all else goes away; it is conceptually impossible not to have a void between the walls and not to have anything else there either. So much the worse, says the reifier, for the

combinatorial principle, which claims that existential statements about distinct things are independent.<sup>10</sup>

(3) We could reify absences reductively. We could identify absences with comparatively uncontroversial objects that, as others would say, are somehow associated with those absences. For instance, we could identify a hole with the hole-lining that, as we'd normally say, immediately surrounds the hole. (Strange to say, some holes are made of cheese and some of limestone! Strange to say, no holes are exactly where we would have thought they were!) Or we could identify an absence with a bit of unoccupied spacetime, if we were not such uncompromising combinatorialists as to countenance an absence of spacetime itself. One way or another, we can cook up ersatz absences to serve as relata of the causal relation—though surely they will seem to be the wrong relata, since we don't really think of these ersatz absences as having the same effects (or causes) as the absences they stand in for.<sup>11</sup> We might, for instance, imitate the identification of holes and hole-linings on a grander scale. Take the most inclusive void of all; and take the mereological fusion of all objects of whatever kind. On the principle of identifying hole with hole-lining, and void with surrounding objects, we might identify this greatest void with the greatest object.<sup>12</sup>

(4) The best response is to concede that a void is nothing at all, and that a lesser absence is nothing relevant at all and therefore cannot furnish causal relata. Yet absences can be causes and effects. So I insist, *contra* Menzies, that causation cannot always be the bearing of a causal relation. No theory of the causal relation, neither Menzies's theory nor any other, can be the whole story of causation.

The intrinsic character of causation is not our present problem. I do indeed fear that the intrinsic character of causation is more a hasty generalization than an *a priori* desideratum.<sup>13</sup> But even if we struck the intrinsic character of causation off our list of folk platitudes, we'd still be trying to characterize the causal relation, so we'd still be in trouble. *Any* relation needs relata, whether it is intrinsic or not. So the problem of missing relata hits any relational analysis of causation.<sup>14</sup>

But does any analysis escape the problem of missing relata?—Yes; a counterfactual analysis escapes. We do not have to reify the void in order to ask what would have happened if the void had not been there. The void causes death to one who is cast into it because if, instead, he had been surrounded by suitable objects, he would not have died. (Here we must assume that if the victim had not been surrounded by the void, he would instead have been surrounded by the life-sustaining objects that normally surround us—not by liquid nitrogen, or clouds of nerve gas, or a hail of bullets.) Likewise for lesser absences. If the cause is an absence, then to suppose away the cause counterfactually is not to attend to some remarkable entity and suppose

that it does *not* exist. Rather, we need only suppose that some *unremarkable* entity *does* exist. Absences are spooky things, and we'd do best not to take them seriously. But absences of absences are no problem.

Note well that in defending a counterfactual analysis, I am not claiming that all causation consists in a relation of counterfactual dependence between (distinct) events. That theory would not escape the problem of missing relata. A relation of counterfactual dependence is still a relation, a relation still needs relata, and absences still fail to provide the needed relata. The counterfactual analysis escapes the problem because, when the relata go missing, it can do without any causal relation at all.

## 6 Menzies's Analysis Retargeted

So far, I have been arguing that Menzies is not entirely right. But in fact I think that a large part of what he says is right, can be separated from the parts that are wrong, and can be accepted even by one who favors a counterfactual analysis. That will be our business for the rest of this essay. Menzies has not given us a fully general analysis of causation, but he has given us something. I think he has given us the right analysis of the wrong analysandum. Let us introduce a name for that which Menzies's functionalist analysis succeeds in analyzing. Let us call it *biff*. That word enjoys some unofficial currency among those who conceive of causation much as Menzies does; so let it be our word for the kind of causation that fits their conception, even if we accept as we should that this is not the only kind of causation.

A theory built around Menzies's three "crucial platitudes" specifies a functional role for a relation: an intrinsic relation between distinct events that is typically, but perhaps not invariably, associated with a probabilistic version of counterfactual dependence. *Biff* is defined to be the occupant of this functional role, if such there be. There is the actual occupant of the biff-role, unless we are badly wrong about the ways of our world. Other possible worlds, some of them, might have different relations occupying the biff-role. In case of imperfect or nonunique occupation of the biff-role, we might resort to semantic satisficing in the ways already considered.

What sort of relation might biff be? We can echo much of what Menzies says, overlooking that he says it not about biff but about causation.<sup>15</sup> Biff—the actual occupant—might, or it might not, be some relation well known to physics. It might, for instance, be force. (More precisely, it might be the relation of exerting a force upon.) Or, taking up David Fair's suggestion about the actual nature of causation,<sup>16</sup> it might be a relation of transfer of energy or momentum. It might be a Humean-supervenient relation. Or it might be a relation posited by some

anti-Humean metaphysic of nomological necessity.<sup>17</sup> It might be a perfectly natural relation, or it might be more or less disjunctive. Myself, I'd like to think that the actual occupant of the biff-role is Humean-supervenient, physical, and at least fairly natural; but nothing else I shall say here is premised on that hope.

Biff might have been one of the relations posited by bygone physical theories that have turned out to be false in our actual world, but are true in various other worlds. Some worlds obey Aristotelian dynamics; the occupant of the biff-role in such a world might be a kind of "force," if we may call it that, which is proportional to velocity rather than acceleration. Or the occupant of the biff-role might have been a relation of transfer of impetus, where impetus is rather like inertia except that it fades away spontaneously, like the heat in a red-hot poker.<sup>18</sup> Or biff might have been some otherworldly physical relation unlike anything that has ever crossed our minds.

Or biff might have been something even stranger. Consider a possible world where occasionalism is true: God is a third party to every causal relationship between events in nature. Then the best available candidate to occupy the biff-role would be an imperfect occupant of the role and an imperfect deserver of the name. It would be a relation between events that did not require accompaniment by anything else in nature but did require accompaniment by God; so it would not be, strictly speaking, an intrinsic relation. Yet we could reasonably judge that this imperfect candidate deserved the name quite well enough.

## 7 Varieties of Causation

Causation by absence is not an instance of biff. Nevertheless it can be described in terms of biff. If you were cast into the deadly void, the absences that would kill you would be absences of biff; because, if instead you were surrounded by suitable objects, events involving those objects would stand in the biff-relation to the events that would constitute your continuing life. Equivalently: Events involving those objects would prevent the event of your death by standing in the biff-relation to events incompatible with it.

Beginning with biff itself, we can define several varieties of causation.

- (1) Event  $c$  directly causes event  $e$  iff  $c$  stands to  $e$  in the relation that occupies the biff-role. For short: iff  $c$  biffs  $e$ .
- (2) The absence of any event of kind  $C$  directly causes event  $e$  iff, had there been an event  $c$  of kind  $C$ ,  $c$  would or might have biffed some event  $d$  incompatible with event  $e$ .

- (3) Event  $c$  directly causes the absence of any event of kind  $E$  iff  $c$  biffs some event  $d$  incompatible with any event of kind  $E$ .
- (4) The absence of any event of kind  $C$  directly causes the absence of any event of kind  $E$  iff, had there been an event  $c$  of kind  $C$ ,  $c$  would or might have biffed some event  $e$  of kind  $E$ .<sup>19</sup>

But there are also cases of *indirect* causation: An event (or absence)  $c$  causes an event (or absence) that then causes another event (or absence)  $e$ . Or  $c$  may cause both an event and an absence, which then jointly cause  $e$ . Or there may be a chain with three steps, or four, or any number. Cases of indirect causation need not reduce to direct causation. Take, for example, a case of causation by double prevention: Event  $c$  causes the absence of any event of kind  $D$ , which absence in turn causes event  $e$ .<sup>20</sup> It does not follow, and it may be false, that  $c$  biffs  $e$ . Even if biff itself is intrinsic, the causal relation of  $c$  to  $e$  in cases of double prevention is sometimes extrinsic.<sup>21</sup>

So the functional analysis of biff affords a basis for defining many varieties of causation. All the varieties there could possibly be? We have no assurance of that. Maybe some possible worlds have no occupant at all of the biff-role, not even an imperfect occupant; and maybe in some such worlds the actual occupant of the biff-role also is nowhere to be found. And when we depart counterfactually from such a biffless world, taking care to avoid gratuitous differences from our starting point, presumably the counterfactual situations we reach will be equally biffless. Yet might there not be some sort of causation in a biffless world? Maybe all the causal relations of events in such a world are thoroughly extrinsic, far more so than in the case of the occasionalist world we imagined before. The “intuition” of the intrinsic character of causation may indeed be right for one basic variety of causation in the actual world, but it is by no means given a priori.

## 8 Ambiguity? Disjunction?

What shall we conclude from this proliferation of different varieties of causation? Has it turned out that we have not one concept of causation but many, so that many different analyses are required to capture the many different senses of the word “cause”? Or is it rather that our one concept of causation is a radically disjunctive concept, so that a correct analysis must consist in a long list of alternative cases? We’re in trouble either way.

If causation in a biffless world is possible, that is a problem equally for both hypotheses. On the many-concepts hypothesis, the problem is that one concept of

causation (at least) will never be reached by our chain of definitions starting with the functional definition of biff. On the disjunctive-concept hypothesis, the problem is that one disjunct (at least) will never be reached. To complete an inventory of senses, or to complete the disjunctive analysis of the single sense, it seems that we must find some different starting point.

Another problem for the many-concept hypothesis is that it requires distinctions in our thinking that sometimes we do not make, need not make, and are in no position to make. If one event directly causes another, for instance, that is causation in one sense; whereas if one event causes another indirectly, in a case of double prevention (or in some still more indirect case) that is causation in a different sense. But when we neither know nor care whether the causation we have in mind is direct or indirect, what concept of causation are we employing then?

(Example: The frightened passenger pulls the cord, knowing that this will cause the train to stop. Being moderately well informed, he knows that pulling the cord opens a valve connecting a reservoir to the outside air. The changed pressure in the reservoir changes the balance of forces on the brake shoes, thereby applying the brakes and stopping the train. But the passenger doesn't know whether this train is fitted with air brakes or vacuum brakes. If air brakes, then the air in the reservoir is normally above atmospheric pressure; so opening the valve lowers the pressure and removes a force, and so the stopping of the train is a case of double prevention. If vacuum brakes, then the air in the reservoir is normally below atmospheric pressure; so opening the valve raises the pressure and applies a force, and so the stopping of the train is a case of direct causation. But so long as he can cause the train to stop, it's all the same to the passenger what kind of causation it is.)

The disjunctive-concept hypothesis now seems better. But it faces an urgent, if not absolutely compulsory, question. Why do we disjoin exactly these disjuncts? Why is the disjunction of just this long list of alternatives anything more than a miscellaneous gerrymander? What makes it a natural kind?

It is as if we came upon some people who had a peculiar taxonomy for birds. They group together a kind that includes swans, but not ducks or geese; eagles and hawks, but not vultures; magpies and crows, but not ravens or currawongs or mudlarks, and indeed no other birds at all. We would be entitled to ask why just these birds are included, and it would not be good enough just to say that all classes are equally classes, and that these people happen to have picked out this class.

The many-concepts hypothesis and the disjunctive-concept hypothesis are both unsatisfactory. Yet if we analyze causation by starting with the functional analysis of biff, and going on to define other varieties of causation one by one in terms of biff, that is the choice we come to.

## 9 Conclusion

I think we are aiming our answers at the wrong question. Menzies went wrong when he took the functional definition of biff to be the whole of a conceptual analysis of causation. We still go wrong if we take it to be even a first step toward conceptual analysis. We should look elsewhere for a conceptual analysis. And we should look elsewhere for a question to which the invocation of biff affords a satisfactory answer.

What is causation? As a matter of analytic necessity, across all possible worlds, what is the unified necessary and sufficient condition for causation?—It is somehow a matter of counterfactual dependence of events (or absences) on other events (or absences).

What is causation? As a matter of contingent fact, what is the feature of this world, and of other possible worlds sufficiently like it, on which the truth values of causal ascriptions supervene?—It is biff: the pattern of relatedness of events to one another by the relation that is the actual occupant of the biff-role. Biff is literally the basic kind of causation, in this world anyway: the basis on which other varieties of causation supervene.

Two different answers to two different questions. They are not in competition. I conjecture that both are right.

If biff is offered not as conceptual analysis but as a basis for supervenience, then it matters little if the varieties of causation, when described in terms of biff, are many and diverse. Unifying the miscellany is a job for conceptual analysis. And if biff is offered as a supervenience basis for causation as it takes place here in our world, then the possibility that some other variety of causation takes place in biffless worlds remote from actuality is no cause for alarm.

Let me say more fully what I have in mind. Doubtless all will agree that the visual qualities of dot-matrix pictures—say, the quality of looking cluttered—supervene on the arrangement of light and dark pixels, at least if we restrict our attention to black and white pictures with maximum contrast. This means that no two possible pictures, at least no two that both fall within our restricted class, differ in their visual qualities without also differing in their arrangement of pixels. If one looks cluttered and the other doesn't, they cannot be alike pixel for pixel. Likewise, the thesis under consideration says that no two possible worlds, or at least no two that fall within a certain restricted class, differ in respect of the truth values of causal ascriptions without also differing in their biff-relations. At least one biff-related pair of events in one world must fail to correspond to any biff-related pair in the other world, and that is what makes the causal difference.

The narrower is the restricted class of worlds within which all causation is said to supervene on biff, the weaker and safer and less interesting our thesis will be. How shall we strike the balance between safety and interest? Since we are especially interested in causation as it takes place in our actual world, the actual world had better fall within the restricted class. But our supervenience thesis, if restricted to one single world, would be utterly trivial; and besides, our interest extends at least to worlds that only narrowly escape being actual. I propose that the restricted class should consist of exactly those worlds that satisfy two conditions. (1) They are worlds where the relation that occupies the biff-role is the same relation that occupies that role in actuality. And (2) they are worlds where the laws of nature are the same as the actual laws of nature.

Condition (2) is not motivated just by caution. Remember our starting point: when the void, or some lesser absence, causes an effect, that is because of the absence of what it takes to prevent that effect (nothing counteracts the vapor pressure that would cause warm blood to boil, and so on). At any rate, that is how causation by absences works under the actual laws of nature, or so we think. But might it not be otherwise under different laws? Suppose there were a fundamental law that said that a certain spell would turn a prince into a frog iff that prince was within a mile of the edge of a void. Under such a law, a void could cause a transmogrification in a way that had nothing to do with absence of biff. There could be such a law—why not? But it is irrelevant to capturing our opinions about how causation works in actuality. So we unabashedly bar the monster: We stipulate that a world with such a law falls outside the range to which our actuality-centered supervenience thesis is meant to apply. And for good measure, but perhaps with needless caution, we likewise stipulate that all worlds that depart from the laws of actuality fall outside the range.

It is because our supervenience thesis is restricted, and because the restriction makes reference to actuality, that our thesis is contingent. Had we started from some different possible world, and restricted according to the same two conditions, we would have restricted the thesis to a different class of worlds. The thesis restricted to worlds that satisfy our conditions relative to this world may be true, but the thesis restricted to worlds that satisfy the same conditions relative to some other world may be false. Or, if we are unlucky, vice versa.

We saw how causation by absences, at least as it takes place in our actual world, could be defined piecemeal using various counterfactuals about biff. So if all causation is to supervene on biff, these counterfactuals about biff must supervene on biff. That is plausible enough. For when we depart counterfactually from a given world, we make no gratuitous changes. Except insofar as the supposition we are making requires differences, the character of the given world carries over into the counter-

factual situation. In particular, the laws governing biff tend to carry over.<sup>22</sup> Return, one last time, to the victim cast into the void (or into something as much like a void as the actual laws allow). If instead he were surrounded by suitable objects, those objects would conform to the laws that actually govern biff. If not, that would be a gratuitous difference between the counterfactual situation and the actual world.

## Notes

1. Here I follow the lead of Martin (1996): “it seems that the void has . . . terrible causal powers” (p. 62). Martin later says that voids are “causally *relevant* but not causally *operative*” (p. 64), but I do not know what he means by this.

2. On combinatorial principles see my (1986e), pp. 86–92; and Armstrong (1989).

3. Menzies (1996) and (1999). Menzies might better have suggested not that causation is an intrinsic relation, but rather that being a causal chain is an intrinsic property. For present purposes, we may leave this correction unmentioned.

4. Here I have combined definitions of intrinsic relations taken from my (1983b), n. 16; and from Langton and Lewis (1998). These definitions differ, but can be expected to pick out the same class of relations.

5. See my (1973a) and (1986b).

6. See my “Causation as Influence,” chapter 3 in this volume.

7. Menzies (1996). The Canberra plan is modeled on analytic functionalism in the philosophy of mind, and on Carnap’s proposal for defining analyticity in a theoretical language. See Carnap (1963).

8. See Bedard (1993).

9. It will be what I called a “diagonalized sense” in my (1970).

10. Casati and Varzi (1994) defend nonreductive reification of holes: Holes are immaterial bodies that depend for their existence on the arrangement of matter. Martin, in his (1996), is probably best classed as a nonreductive reifier, despite his emphatic warnings that absences are not things. (His conflicting suggestion that they are “localized states of the universe” would seem to be retracted by his denial that they are natural properties of things [p. 58]. The universe would seem to be a thing, states would seem to be properties, and properties of local emptiness would seem to be not unnatural.)

11. Lewis and Lewis (1970) is, for the most part, a dialogue between a reductive and a nonreductive reifier. Neither one notices the option of paraphrasing hole-statements in terms of quantification over ersatz holes without also claiming that holes are identical to ersatz holes. Frank Jackson calls attention to that option in his (1977), p. 132.

12. Or, since the greatest object has the property of totality—of being all there is—we might identify the greatest void with the having by that object of the property of totality. This ersatz void is as wrong as it can be in its effects (and causes): It causes what objects cause, not what the void unassisted by objects causes. And it is as wrong as it can be in its location, being exactly where the void isn’t. It violates a combinatorial principle, since it cannot possibly coexist with any extra object. For that very reason it is of doubtful ontological status, being a having of a merely extrinsic property.

D. M. Armstrong uses havings of totality as truthmakers for negative existential statements, *inter alia* in his (1997), chapter 13. But he never asks these totality states of affairs to serve as ersatz absences.

13. See “Causation as Influence,” chapter 3 in this volume.

14. How about an analysis in terms of a relation between propositions, where, in the case of causation by or of an absence, one of the related propositions is a negative existential?—Not problematic; but not what I’d call a relational analysis of causation. We don’t want to say that a cause- or effect-describing proposition is *itself* the cause or effect.

15. Menzies (1999).

16. Fair (1979).

17. It might for instance be the relation of instantiating a law, where a law is taken to be at once a state of affairs involving universals and a universal in its own right. See Heathcote and Armstrong (1991).

18. See Butterfield (1957), pp. 8–13.

19. The definitions of cases (2)–(4) resemble, but differ from, the definitions given in Fair (1979), p. 247, and also the somewhat different definitions given in the final section of Dowe (1999). Note that these counterfactual definitions are not the same as the counterfactuals about events and absences that would appear in a counterfactual analysis of causation.

20. See Ned Hall, “Two Concepts of Causation,” chapter 9 in this volume; and McDermott (1995a).

21. As is noted in McDermott (1995a) in connection with his example of Nixon, Haig, and Joe Blow’s breakfast. See also the billiard-ball example in “Causation as Influence,” p. 84 of this volume.

22. Unless violating a law yields the least departure overall. See Lewis (1979a).