Events and Times: A Case Study in Means-Ends Metaphysics

Christopher Hitchcock

Division of Humanities and Social Sciences

MC 101-40

California Institute of Technology

Pasadena CA 91125

USA

Events and Times: A Case Study in Means-Ends Metaphysics

1. Introduction

This is an essay in metaphysics, scientific methodology, and metametaphysics. In metaphysics, it addresses an old debate about the nature of events, and their relation to time. In methodology, it addresses considerably newer issues about how to construct appropriate causal models. In metametaphysics, it argues that problems of the first sort, and perhaps metaphysical problems more generally, can fruitfully be recast as problems of the second sort. This is not intended to be a grand vision of what all metaphysics must aspire to, but it is intended to be an attractive picture of at least some parts of metaphysics -- one that builds on an important historical tradition.

2. Historical Background

I begin by briefly retracing a particular trajectory through the history of metaphysics.[1] For the Continental rationalists such as Descartes and his followers, metaphysics, or 'first philosophy', was distinguished from natural philosophy by its subject matter, but not by its methods. While natural philosophy concerned the observable world, and encompassed what we now recognize as the beginnings of modern empirical science, metaphysics

[1] Here I lean heavily on the historical work of Michael Friedman. See for example, the short overview in Friedman 2001, chapters I and II.

concerned unobservable entities such as God, substance, and the soul.[2] Since the rationalists emphasized the application of the intellect in inquiry, rather than sense experience, this difference between the subject matters of natural philosophy and metaphysics did not entail a substantial difference in the type of epistemic access we have to the two different domains. God, substance, and the soul were no more epistemically problematic than the unobservable tiny corpuscles that populated Cartesian natural philosophy.[3]

Kant's *Critique of Pure Reason* attempted to fundamentally transform the nature of metaphysics. For Kant, metaphysics was not the study of some particular class of entities. Such entities might exist, but if so, they are epistemically inaccessible to us. Rather, for Kant, metaphysics concerned the necessary pre-conditions for empirical knowledge. For Kant, the mind had to impose a certain structure upon the world in order

---

[2] The status of the 'soul', or mind in the rationalist period is somewhat tricky. For Aristotle, the soul was what animated a living being ('animation' being etymologically related to 'anima'), and was very much a part of the natural world. The traditional understanding of Descartes has him rejecting this view, attributing the soul to human beings only, and moving it from the realm of natural philosophy to metaphysics. The truth, however, is considerably more complicated, and the status of the mind as an object of inquiry remained in flux throughout the Seventeenth Century. See, e.g. Manning (forthcoming) for further discussion.

[3] See Laurie Paul's contribution to this volume for a contemporary version of this picture of metaphysics.

for it to be comprehensible to us, and metaphysics was the study of this imposed structure. Kant carved up the terrain somewhat differently than his predecessors. Metaphysics proper concerned 'categories' like causation and substance that were necessary for having propositional knowledge of the world. The 'transcendental aesthetic' included the 'intuitions' of (Euclidean) space and time, which were essential pre-requisites for having coherent perceptual experience. Other concepts, like freedom and God, found their place as pre-requisites for practical reason.

This transformation of the subject matter of metaphysics brought with it a fundamental change in metaphysical methodology. Knowledge of the world was essentially empirical knowledge, as generated by the empirical sciences. Knowledge of metaphysical truths, such as the law of causality, was arrived at through the 'transcendental deduction'. The deduction is 'transcendental' in the sense that it does not show that, e.g. the law of causality is *a priori* necessary, but only that it is necessary *if there is to be empirical knowledge at all.* For Kant, then, conceiving of the world in certain metaphysical categories was a *means* to achieving certain epistemological *ends*.

Kant's reconfiguration of metaphysics was echoed in the work of many of the Logical Empiricists. I will consider two examples: Rudolf Carnap and Karl Popper. I want to focus here on two elements of Carnap's philosophy. The first is his articulation of the project of 'explication' as a distinctive philosophical project (Carnap 1950a). Explication differs from the more familiar activity of 'conceptual analysis' as defended, for example, by Strawson (1963) and more recently by Jackson (1998). In conceptual analysis, a concept such as 'causation' or 'simultaneity' is seen as being 'grasped' by competent speakers of English. The contours of the concept can be explored by eliciting the judgments of competent speakers about how the term is to be used, typically in some range of hypothetical cases. These judgments are often referred to as 'intuitions', and serve as the evidential basis for philosophical theory construction. Carnap thought that ordinary concepts such as 'causation' are imprecise, ambiguous, and incompletely articulated. To explicate such a concept is to replace it with a concept that is clear and precise, and which captures some prominent aspect(s) of the usage of the original term. It is sometimes forgotten, however, that for Carnap, explication was not an end in itself. Rather, the goal was to produce a concept that was well designed for some specific task, typically the articulation of claims in empirical science. An explicated concept was, for Carnap, like a well-designed scientific instrument. The bread knife in my kitchen is fine for slicing bread, but it can't be used to prepare cross sections of a cell for viewing under a microscope. Likewise, our ordinary notion of simultaneity works well enough for purposes of scheduling a meeting, but not for doing fundamental physics. (See, e.g. Carnap's response to Strawson in Carnap 1963).

The second feature of Carnap's view that I want to draw attention to is his distinction between 'internal' and 'external' questions, as articulated in Carnap (1950b). There he is particularly concerned with ontological questions that arise in fields like mathematics and formal semantics. Carnap maintains that any type of inquiry must be carried out within some linguistic framework. The adoption of a linguistic framework is a pre-condition for making any kind of meaningful claim. Internal questions are those that can be answered using the resources of a given linguistic framework. Their answers are determined by the rules of the language, and hence are analytic. For example, in the linguistic framework of Peano arithmetic, the internal question of whether numbers exist has a clear, positive, answer. External questions concern which linguistic framework we should adopt. Here, we cannot appeal to the rules of any one linguistic framework. External questions, by their nature, are framework-independent. Carnap thought that external questions could only be answered on pragmatic grounds, e.g. by appeal to the fruitfulness of inquiry pursued using a given framework, but Carnap himself provided few details about how to go about answering such questions. A metaphysical question, such as ‘do numbers *really* exist?’, is meaningless unless it can be re-construed as either an internal question (e.g. about whether numbers exist in the framework of Peano arithmetic) or an external question (e.g. about the fruitfulness of Peano arithmetic as a linguistic framework). The take-home message for my project is that Carnap thought that some metaphysical questions, especially ontological questions, were properly recast as questions about the utility of adopting a certain kind of language for a specific field of inquiry.

The inclusion of Popper in my discussion may seem surprising: a positive conception of metaphysics does not play a central role in Popper’s philosophy. A negative conception does play a central role: the demarcation of empirical science from other types of inquiry is a central concern of Popper’s, so insofar as metaphysics is distinct from empirical science, Popper’s conception of science provides a kind of negative characterization of metaphysics. But the feature of Popper’s work that I want to emphasize here is his repeated use of a strategy of translating metaphysical theses into methodological rules. For example, concerning the ‘principle of causality’ he writes:

> I shall…propose a methodological rule which corresponds so closely to the 'principle of causality' that the latter might be regarded as its metaphysical version. It is the simple rule that we are not to abandon the search for universal laws and for a coherent theoretical system, nor ever give up our attempts to explain causally any kind of event we can describe. (1959, 39)

Concerning the principle of the uniformity of nature, he writes:

> Let us suppose that the sun will not rise tomorrow…Existing theories would presumably require to be drastically revised [sic]. But the revised theories would not merely have to account for the new state of affairs: *our older experiences would also have to be derivable from them*. From the methodological point of view one sees that the principle of the uniformity of nature is here replaced by the postulate of the *invariance of natural laws* with respect to both space and time…[T]he 'principle of the uniformity of nature' can…be regarded as a metaphysical interpretation of a methodological rule… (1959, 250 – 1)

Popper maintains that his methodological prescriptions have the status of *conventions*. These methodological rules define a certain kind of activity (which, in his English translation of 1959, he sometimes calls by the name 'experience'), in the way that the rules of chess define a particular game. The program of his *Logic of Scientific Discovery* is to articulate these methodological rules, and to demonstrate that a system of inquiry based upon these rules would have a variety of desirable features.

3. Means-ends Metaphysics

What Kant, Carnap, and Popper have in common is their strategy for recasting issues in metaphysics as issues about the utility of certain conceptual structures for achieving broadly epistemological goals. My own approach is considerably less ambitious than the programs of these three luminaries. In Kant, the program was to justify the framework of traditional Aristotelian metaphysics by showing that it is a means to achieving knowledge. Carnap intended his notion of a linguistic framework to be a pre-requisite to any inquiry. Popper sought a conception of the ideal inquiry of science. The scope of my project is much more modest.

There is, however, one dimension along which I aspire to more than Kant, Carnap, or Popper was able to achieve. Kant's transcendental deduction is no longer considered cogent; similarly Popper's views have been severely criticized. Carnap, as we have seen, offered little by way of explanation of how one would actually justify the use of a particular linguistic framework. By contrast, I intend for my own 'transcendental deduction' to be much more rigorous. I take as my model the approach to epistemology that Schulte (1999) calls 'means-ends epistemology'. The methodology of means-ends epistemology is to justify inferential rules by articulating precise epistemic goals, and demonstrating that the rules in question will satisfy those goals. One illustration is Reichenbach's (1949) pragmatic justification for his 'straight rule' of induction. Suppose that one has observed *n* tosses of a coin, and *m* of them have resulted in heads. What should one infer about the probability of heads? Reichenbach posited as the goal of inquiry convergence to the true probability (if it exists) in the limit. According to

Reichenbach, the probability of heads is just the limiting relative frequency of heads (if the limit exists). One can be assured of converging to the correct probability (again assuming that it exists) by positing that the probability is *m/n* after observing *m* heads in *n* tosses.[4]

We have arrived, at last, at a point where I can articulate my own proposal. My approach comes closest to Popper's. The idea is to combine the Popperian strategy of transforming metaphysical theses into methodological rules with a rigorous means-ends approach to justifying those methodological rules, resulting in a means-ends metaphysics. While I do not claim that all metaphysics can be done in this way, I think that it represents an attractive blueprint for tackling a number of metaphysical problems.

4. Events and Times

The particular metaphysical issues that I will use to illustrate means-ends metaphysics concern the relationship between events and times. One such issue is whether the time at which an event occurs is *essential* to that event. For example, if a meeting is originally scheduled for Monday at noon, and then re-scheduled for Tuesday at noon, is the meeting that occurs on Tuesday at noon the very same meeting that would have occurred on Monday? That is, was the meeting *postponed*, strictly speaking, or was the original meeting canceled and a *different* meeting scheduled for Tuesday? Lombard (1986) for

[4] While the straight rule is sufficient for achieving the goal of convergence to the correct probability (if it exists), it is not necessary, as a number of commentators have pointed out.

example, argues that the time at which an event occurs *is* an essential property of that event;[5] Lewis (1986) argues for the opposite position.

(Note that it is often awkward to talk, at least informally, in a way that does not seem to presuppose an answer to this question. For instance, we might ask whether the Department Chair's vacation schedule affected the time at which the meeting was held. Such a question seems to presuppose that the meeting could have been held at a different time; otherwise we should ask whether the Chair's schedule affected *which* meeting was held. To avoid circumlocutions, I will continue to talk informally using our ordinary expressions. Such talk should not be read as begging the question on this issue.)

How shall we attempt to address this issue? Consulting our intuitions will not take us very far: few of us have strong pre-theoretic intuitions about such arcane matters. Lewis (1986) attempts to defend his view by appeal to the needs of a counterfactual theory of causation. Note that this is already adopting a means-ends approach to the metaphysics of events. Events enter philosophical discussion at a number of points. For

[5] This view is also sometimes attributed to Kim (1973a). Kim takes the time at which an event occurs to be a distinguishing feature of that event. But it seems that Kim's purpose is to distinguish an event from other events that actually occur. For example, if my doorbell rings at noon, and then again at one o'clock, the two ringings are distinguished by their time of occurrence. Kim does not explicitly propose that the time at which an event occurs should also serve to distinguish it from merely possible events. So it is unclear whether he intends that the time at which an event occurs is an essential property of that event.

Lewis, the point of entry is the analysis of causation. Davidson (1967a, 1967b) offers a quite different rationale for including events in our ontology: capturing inferences involving adverbial modification. For example, we can infer from ‘Jones ran quickly’ to ‘Jones ran’. It is hard to see how to do this if we parse ‘Jones ran quickly’ as asserting that a certain property held of Jones. Davidson argues that in order to capture this inference, we need to introduce an entity, a running, or perhaps a running by Jones, that has the property of being quick. The needs of a theory of adverbial modification may be quite different from those of a theory of causation. We will follow Lewis in approaching these issues through the lens of causation.

According to a simple counterfactual theory of causation, event *C* is a cause of event *E* just in case (i) *C* and *E* both occur, (ii) *C* and *E* are suitably distinct (e.g. not related logically or by spatiotemporal overlap), and (iii) the following counterfactual is true: ‘if *C* had not occurred, *E* would not have occurred’.[6] Clause (ii) is necessary in light of problems raised by Kim (1973b). Without clause (ii), we would be forced to say that Jones’ saying ‘hello’ caused him to say ‘hello’ loudly (if he had not said it, he wouldn’t have said it loudly), or that his pronouncing ‘ell’ caused him to say ‘hello’ (for if he had not pronounced ‘ell’, he would have said ‘ho’ instead of ‘hello’). Lewis (1986) offers a detailed account of the relevant notion of distinctness. This simple theory of causation fails for well-known reasons involving redundant causation, but we will put these problems aside.

---

[6] Lewis (1973) offers this as an account of *causal dependence*, which is sufficient, but not necessary for causation.

A counterfactual theory of causation connects our question about event essences to a different question: if *C* affects the time at which *E* (or a numerically distinct, but similar event) occurs, does *C* count as a cause of *E*? Let us call such an event *C* a *hastener* or *delayer* for *E*. Bennett (1987) illustrates this point with the following example. There have been heavy rains in a certain forest during the month of March. Throughout April and May, the forest is exposed to a number of lightning strikes. Since the forest is still damp in April, it does not catch fire then. By May, the forest has dried, and the lightning strikes ignite a fire. Did the March rains cause the fire? If the rains had not occurred, there would have been a fire in April, but not one in May (since there would have been no fuel left to burn). The March rains were a delayer of the forest fire. In order to know whether the March rains *caused* the fire, we need to know whether the fire would have occurred without the rains. And in order to know this, we need to know whether the fire that would have occurred in April would have been the *same* fire. Lewis argues that if the time at which an event occurred is essential to it, then any event that influences the time at which some other event occurs will qualify as a cause of that event. According to Lewis, this would lead to an unacceptable proliferation of causes.

Our question about the essences of events need not be tied directly to the causal status of hasteners or delayers. Paul (1998) suggests that we amend our counterfactual theory of causation to read that *C* is a cause of *E* if (i) they both occur, (ii) are distinct, and (iii) if *C* had not occurred, then either *E* would not have occurred, or else *E* would have occurred at a different time. This proposal would settle the causal status of hasteners and delayers without having to settle the question about event essences. Nonetheless, I think that the two problems remain closely connected. Both concern the question of whether differences in the time of occurrence of an event have a kind of special status that distinguishes them from other kinds of differences in the way in which an event can occur.

5. Causal Models

The past two decades have seen tremendous progress in the development of formal tools for representing causal systems, and for facilitating causal inferences. Among philosophers, the most influential treatments have been those of Pearl (2000, 2009), and Spirtes, Glymour and Scheines (2000). They develop a variety of tools for representing causal systems graphically, including structural equation models and causal Bayes nets. These models can be used, for example, to infer qualitative causal structure and to estimate causally significant quantities using only information about probabilistic correlations. The successes of these approaches are clearly sufficient to meet Carnap's standard of fruitfulness.

I will here make use of only the simplest kind of structural equation models. I will illustrate the use of these models with a simple example. One night, I turn on the gas grill in my backyard, and grill a steak. (Apologies to vegetarian readers; I hope only that they will find the example no more objectionable than standard examples involving assassination.) To light the grill, I have to turn a knob, which controls the gas, and press a button, which ignites a spark. I then put the steak on the grill, and it is cooked in about ten minutes. To represent this little system, I begin with some *variables*: *Igniter, Gas, Lights, Put steak on,* and *Steak cooked. Igniter* takes the value '1' if I press the igniter button, '0' otherwise; *Gas* takes the value '1' if I turn the gas knob, '0' otherwise; and so on for the other variables. I can then represent the structure of the system by expressing the value of each variable in the system as a function of its immediate causal predecessors:[7]

$$Lights = Igniter \times Gas$$

$$Steak\ cooked = Lights \times Put\ steak\ on$$

The first equation tells us that the grill lights just in case the igniter is pressed and the gas knob is turned; the second equation tells us that the steak will get cooked just in case the grill lights, and the steak is put on the grill. The variables *Igniter, Gas,* and *Put steak on* are *exogenous*: their values are not determined by the other variables in the system. In fact, these variables all took the value '1'. The variables *Lights* and *Steak cooked*, which appear on the left hand side of the two equations, are *endogenous.* The values of the exogenous variables, together with the equations for the endogenous variables, entail exactly one set of values for all of the variables in the system. In our example, all of the variables take the value '1'. The causal structure can be represented graphically as in figure 1. An arrow from one variable to another indicates that the former variable figures on the right hand side of the equation for the latter variable.

[figure 1 approximately here]

[7] 'Immediacy' here is relative to the set of variables chosen, rather than absolute.

As described, all of the variables are binary: they take the value '0' or '1'. The value '1', intuitively, corresponds to the occurrence of some event, while the value '0' corresponds to the absence of the event in question. In fact, however, I can turn the gas knob to different levels: low, medium, high, and points in between. The flame on the grill varies correspondingly, and depending on the level of the flame, the steak might be rare, medium, or well done after ten minutes of cooking. We can represent this by allowing the variables *Gas, Lights,* and *Steak cooked* to take values between '0' and '10'. Our equations can remain same. The first equation, for instance, will tell us that the grill will light with a flame equal to the level of the gas, so long as the igniter is pressed. If the igniter is not pressed, *Lights* will take the value '0'. Here, the different values of the variables correspond to what Lewis (2000) calls different *alterations* of the event in question.[8]

In using *variables* to represent causal relations, we have changed the language that we use to talk about causal relations. While the values of variables might naturally be thought of as *events* in the standard philosophical sense (or as *alterations,* in the sense of Lewis 2000), 'event' is not literally part of the vocabulary now being used to talk about causation. The use of variables instead of events is standard in almost all scientific contexts, and I have argued extensively in a number of other publications (e.g. Hitchcock 1993) that they are more appropriate for many everyday contexts as well.

[8] Lewis is explicitly neutral on whether different alterations of an event constitute different events, or different ways in which the very same event could occur.

The equations given above describe the pattern of dependence that holds for this specific grill on a particular night. They are not laws of nature, and they do not describe causal generalizations. If the gas to my house is turned off[9], perhaps as the result of an earthquake[10], the grill will not light, even if I turn the gas knob and press the igniter. If I wanted to, I could represent this explicitly by adding a variable to the causal model representing whether the gas is working. But since the gas was in fact working on the night in question, it is not necessary to do this, and the equations are correct as they stand.

So far, our equations allow us to draw inferences in both directions of time. For instance, if I learn that the grill has lit and that steak has been put on the grill, I can use the model to infer that the steak will cook. Similarly, if I learn that the steak has cooked, I can infer that the grill was lit and that the steak was put on the grill. The asymmetry of causation enters in the way that causal models are used to represent *interventions* or *counterfactuals*. In the causal modeling literature (e.g. Pearl 2000, 2009, Woodward 2003), the emphasis has been on talk of interventions. Metaphysicians, especially students of Lewis (either literally or figuratively), may be more comfortable with talk of counterfactuals, specifically of what Lewis (1979) calls 'non-backtracking counterfactuals'. The difference is essentially one of temporal perspective: looking forward in time, it is more natural to think in terms of interventions; after the fact,

---

[9] My backyard grill is connected to the main gas line of the house, rather than having a separate gas supply.

[10] The gas supply to my house is equipped with a seismic shut-off valve, which closes the gas line if there is an earthquake, or if the plumber bangs on the pipes too hard.

looking back retrospectively, it is more natural to think in terms of counterfactuals. I will not draw a distinction between the two.

Suppose we want to know what would have happened if the grill had not lit, i.e. if the system were intervened on so as to prevent the grill from lighting. The key idea here is that an intervention involves a variable taking a value through some means that over-rides the normal causal structure. Woodward (2003) thinks about this in terms of some kind of independent and exogenous causal process. For example, if the grill had been filled with water, it would not have lit, regardless of whether the gas was turned on and the igniter pressed. Lewis (1979) thinks of the antecedent of a non-backtracking counterfactual as being brought about by a 'small miracle', a local violation of the normal laws of nature. For present purposes, these amount to essentially the same thing: the variable *Lights* taking the value '0' in some way that overrides the original equation in our causal model.

An intervention that sets the value of *Lights* to '0' is *not* represented by simply plugging the value '0' for *Lights* into our causal model. That would lead to backtracking: we would infer that either the gas knob had not been turned, or that the igniter had not been pressed. Rather, we represent the intervention by *replacing* our original equation for *Lights* with a new one specifying that *Lights* = 0. This has the effect of changing the causal structure to one in which *Gas* and *Igniter* are no longer causes of *Lights*, and for this reason, an intervention is sometimes said to 'break the arrows' into the variable that is intervened on (see figure 2). By replacing the equation for *Lights*, the new hypothetical value of that variable is propagated forward through the causal model, but not backward.

[Figure 2 approximately here]

The effects of interventions are empirically testable, although such tests are defeasible in at least two ways. On the night in question, the grill did in fact light. We cannot go back and intervene to prevent it from doing so. If I want to discover the effect of intervening to prevent the grill from lighting, I must do so using other grills, or the same grill on different occasions. Inferences drawn from such experiments then presuppose that the causal structure of the grill that I actually intervened on is relevantly similar to the one being modeled. Additionally, such inferences presuppose that my actual experimental manipulation has the right kind of causal structure to mimic a true intervention (or Lewisian miracle). If I prevent the grill from lighting by turning the gas knob to 'off', I am not intervening on the variable *Lights*, but rather on the variable *Gas*. But these are just the sorts of caveats that apply to all inferences from experiments.

Even putting aside the empirical successes of causal modeling techniques, this kind of causal model strikes me as an extremely natural representation of our simple causal system. I think it plausibly reflects the kind of representation we employ (even if only tacitly) in our everyday deliberations about how to intervene in the world.

While the formal tools of causal modeling have been well-developed (e.g. in Pearl (2000/2009) and Spirtes, Glymour and Scheines (2000)), and the interpretation of causal models well-understood (through Woodward (2003), as well as the sources mentioned previously), the *methodology* of causal modeling remains under-explored. Specifically, there are a number of questions relating to the 'art' of modeling. How does one construct an appropriate model? What variables should one use? When does one have enough variables? And so on. While Halpern and Pearl (2005) have stressed the freedom of the modeler to represent a causal system in different ways, there are some important constraints on the construction of causal models. I have explored some of these in Hitchcock (2001), Hitchcock (2004), and Halpern and Hitchcock (2010). In the following sections, I offer a further contribution to this discussion.

6. Reformulating the Problem

Within our causal modeling framework, we can give a very precise and tractable reformulation of our metaphysical questions about times and events. Suppose that we wish to model a structure like the one described by Bennett (1987), where we have an event (or several closely related events) that may happen at different times. How shall we model this structure? In particular, what variable(s) should we include in our model to represent the event(s) in question?

There seem to be at least three prima facie plausible candidates for representing the forest fire(s) in Bennett's example. First, we could have a single binary variable, *FF* whose values represent the following possibilities:

FF = 1 if a forest fire occurs at some time in the scenario

FF = 0 if there is no forest fire

Choosing to represent the forest fire in this way would not allow us to distinguish between a forest fire in April, and one in May. But if these possibilities really do correspond to the occurrence of the same event, perhaps there will be no harm in collapsing these two possibilities together into the same value of one variable.

A second possibility would be to have a multiple-valued variable FF_t, whose values represent the different possibilities in the following way:

$FF_t = 0$ if there is no forest fire

$FF_t = 1$ if there is a forest fire in April

$FF_t = 2$ if there is a forest fire in May

Now we can distinguish between a fire in April and a fire in May, since these two possibilities are represented by different values of the variable FF_t.[11]

A third way to represent the situation would be to use two different variables, FF_a and FF_m with the following interpretations:

$FF_a = 1$ if there is a forest fire in April

$FF_a = 0$ if there is no forest fire in April

$FF_m = 1$ if there is a forest fire in May

$FF_m = 0$ if there is no forest fire in May

This choice gives us slightly greater expressive power, since it allows us to represent the occurrence of a forest fire in April *and* a forest fire in May, by letting both variables take

[11] Halpern and Pearl (2005) model Bennett's example using just this kind of variable.

the value 1. We might think, however, that while this choice of representation allows us to represent an additional *logical* possibility, this won't really do us any good, since this possibility is not one that could actually occur. (If there were a forest fire in April, there would be insufficient fuel to sustain a fire in May.)

The problem, then, is whether a forest fire in April and a forest fire in May correspond to the *same value* of the *same variable*, to *different values* of the *same variable*, or to values of *different variables*. I think that our reformulated problem is appropriately seen as the translation of our original cluster of metaphysical problems into the language of causal modeling. Our causal models represent causal relations among *variables*, or perhaps *values* of variables, rather than between events. So the reformulated problem is not worded in exactly the way our original problems were. In this sense, the reformulated problem has the character of a Carnappian explication. But it is, I believe, the closest we can come to translating our original problem into our new, more precise, language.

The question of how best to represent this kind of situation should be compelling even for those who normally disdain metaphysics; for the question can be motivated without any appeal to the notion of events and their essences. If we want to be able to represent the possible causes of forest fires, and to plan interventions to prevent forest fires, we will need to know which representation to use. Putting the point in terms of the framework of means-ends metaphysics, we have a specific end in sight. We want to be able to construct a causal model that will accurately tell us what the effects of various interventions will be.

I will argue that only the third type of representation, which employs two separate variables, will always be able to do that. This is not to say that collapsing the different possibilities in the ways done by the first two approaches will *always* lead to trouble. Arguably, in Bennett's own example, we could successfully model the situation using any of the three strategies introduced above.[12] In the next section, however, I will introduce a variant on Bennett's example where only the third strategy will work.

7. The Careless Camper

Get rid of the rain, and meet a new source of danger for our forest. A character we will call the 'Careless Camper' (or CC for short) has planned a camping trip for the first weekend in May. Of course, if there is a forest fire in April, he will cancel his trip. Who wants to camp out in a burnt-down forest? If he does go camping, being careless, he will leave a fire unattended. The unattended fire will cause a forest fire, unless something prevents it from doing so (such as the absence of fuel from an earlier fire). As it happens, there is a forest fire in April, CC does not go camping, and there is no unattended fire.[13]

[12] This is a bit tricky. If we want to distinguish between the lightning strikes in April and those in May, and have our model reflect that the latter caused the fire while the former did not, then we cannot get by with the first strategy for modeling the forest fire. The second strategy, however, will not cause us problems even in this case. For example, I don't think the use of the second strategy in Halpern and Pearl (2005) undermines their analysis of the example.

[13] This is a variant of an example used in Halpern and Hitchcock (2010).

We want to build a causal model of this little story. To start with, we will include two variables, CC and UF:

$CC = 1$ if CC goes camping in early May

$CC = 0$ if CC does not go camping then

$UF = 1$ if there is an unattended fire in early May

$UF = 0$ if there is not an unattended fire then

Our model will certainly include as an equation:

$$UF = CC$$

This equation tells us that there will be an unattended fire just in case CC goes camping. But how shall we build the possibility of forest fire into our model? We have three candidates from the previous section.

The first option is to use the binary variable FF. What will our equations look like? Well, CC won't go camping if there is a forest fire (in April). And there will be a forest fire (in May) if CC goes camping. So this suggests that our equations should be something like this:

$$CC = 1 - FF$$

$$UF = CC$$

$$FF = UF$$

This system of equations is represented graphically in figure 3. The graph shows a cycle, which suggests that something has gone wrong. Worse, our system of equations has no consistent solution! These equations tell us that there will be a forest fire just in case there is no forest fire. Indeed, the causal model looks like a representation of the sort of self-defeating causal loop we find in the grandfather paradox for time travel.[14] Obviously, we have been crippled by our inability to distinguish between a forest fire in April, which would *prevent* CC from going camping, and a forest fire in May, which would be an *effect* of his going camping. With no way of distinguishing between the two possibilities, we seem forced to say that the forest fire would prevent one of its own causes!

[Figure 3 approximately here]

The second strategy, using the ternary variable FF_t for representing the forest fire looks more promising, since it can distinguish between a fire in April and a fire in May. What will our equations look like? CC will go camping unless there is a fire in April; so CC will be 1 if FF_t is 0 or 2. So we can begin to write our equations as follows:

$CC = 0$ if $FF_t = 1$

$CC = 1$ if $FF_t = 0$ or 2

[14] In fact it is worse. As Lewis (1976) points out, in the case of the grandfather paradox there is a consistent solution: one where the time traveler fails to kill her grandfather.

$UF = CC$

The problem, however, is that there is no satisfactory way to write an equation for FF_t. If we write an equation with FF_t as a (non-trivial) function of CC or UF, that would entail that intervening to make CC go camping, or to create an unattended fire in May (or some combination), would lead to some result other than a forest fire in April. But this is wrong. For instance, if we had intervened to light a fire in early May and leave it unattended, the forest fire still would have occurred in April and not in May. So this suggests that we treat FF_t as an exogenous variable with constant value:

$FF_t = 1.$

This structure is shown graphically in figure 4. The problem now is that we have no way to capture the idea that intervening to prevent a forest fire in April will lead to CC's going camping, an unattended fire, and hence to a forest fire in May. If we set FF_t to a value other than 1, our equations tells us that CC will go camping, and that there will be an unattended fire. But then they stop. They have no way of telling us that an unattended fire will lead to a forest fire in May, since our equation for FF_t does not make it dependent on UF. It seems that we want FF_t both to depend upon UF and not depend upon it.

[Figure 4 approximately here]

It's pretty easy to see where we have gone wrong. Whether or not there is an unattended fire in May makes no difference to whether there is a fire in April, but it could make a difference to whether there is a fire in May. By including these two possibilities as values of the same variable, we have no consistent way to represent the dependence of the forest fire upon the unattended fire.

By contrast, our third strategy, using two separate variables FF_a and FF_m, leads to smooth sailing. We can write our equations as follows:

$$CC = 1 - FF_a$$

$$UF = CC$$

$$FF_m = UF \times (1 - FF_a)$$

The graph is shown in figure 5. The last equation tells us that there will be a forest fire in May just in case there is an unattended fire in May, and there was no forest fire in April. These equations correctly tell us that in the actual case, where there is a fire in April, CC will not go camping, there will be no unattended fire, and there will be no fire in May. They also tell us that if there had been no fire in April, *CC* would have gone camping, he would have left an unattended fire, and there would have been a forest fire in May. Moreover, they correctly tell us that in the actual case, where there was a fire in April, intervening to leave an unattended fire in May would not have resulted in a forest fire in May. Readers can verify that the equations correctly capture all of the correct interventions and counterfactuals.

[Figure 5 approximately here]

Let me end this section with a remark about the way in which the third representation of our little story was motivated. In my exposition, I introduced structural equations models in section 5, and motivated their use by appeal to their various successful applications. I then asked the question of how, in this framework, we should go about choosing the right set of variables to use in modeling the story of the Careless Camper. I argued that only the model using two separate variables, FF_a and FF_m, could capture all of the correct counterfactual and intervention relations. Thinking in Carnappian terms, it appears that the choice of language was motivated by broad appeal to fruitfulness, and means-ends analysis was only invoked to select among several possible dialects of this language. So it looks as though the broad appeal to fruitfulness is doing most of the work, and means-ends analysis is only needed to add the finishing touch.

Although this is indeed how I presented the argument, I think that the structure of the argument is easily re-worked so that means-ends analysis is carrying the full load. The goal is to represent the story of the Careless Camper in a way that correctly captures the correct counterfactuals/intervention relations. I showed that the final structural equation model with the two separate variables FF_a and FF_m satisfies this goal. This can be established without any need to invoke the other various successes of structural equation models. I also showed that two alternate structural equation models, using different variables, were inadequate. What I have not done, of course, is show that *no* other representation (not even one using very different formal or conceptual tools) can be adequate. That would clearly be an impossible task.

8. Why Time is Special

The key to the success of the third strategy for modeling the story of the Careless Camper is that by using two separate variables, FF_a and FF_m, we can capture the key fact that whether there is a fire in May depends *causally* on whether there is a fire in April. This is something that we just can't do if we use only one variable. This insight points us to a much broader moral. When we introduced the simple counterfactual theory of causation in section 4, we added the standard caveat that cause and effect need to be *distinct* events. The worry is that one event may counterfactually depend upon another if the two events are logically or spatiotemporally related in certain ways, and we do not want to count these as cases of causation. There is a fairly obvious extension of this principle: if two (possible) events are mutually *exclusive* on logical and/or spatiotemporal grounds, we do not want to say that one of them *prevents* the other. Consider again the gas knob on my backyard grill. I can set it to low, medium, or high. If I set it to high, then it won't be set to low. But my setting it to high does not *prevent* the knob from being set to low. The relation between the two is not causal, but rather logical and spatiotemporal. (It can't be set to both low and high at one and the same time.) One of the basic methodological rules for the construction of causal models is that possibilities that are mutually exclusive in this way should be represented as different values of the same variable.[15] Otherwise our

---

[15] See e.g. Hithcock (2001) for discussion of some of the reasons for this rule.

causal models will represent logical and/or spatiotemporal relations as though they were causal relations.

What we have seen in the previous section is the importance of the converse methodological rule. If two possibilities are mutually exclusive because one *prevents* the other, we should, in general, represent them as values of *different* variables. If we represent them as values of the same variable, we cannot capture the causal relationship between them. Depending upon what else we want to include in our causal model, this may lead to all sorts of trouble. A great many of the standard philosophical examples involving events that can occur at different times have exactly this character: a forest fire that occurs at one time may prevent a later forest fire; a bottle's breaking at one time may prevent it from breaking at a later time; a person's death at one time may prevent her from dying at a later time; and so on. By contrast, different possible states of an object or system *at the same time* do not stand in relations of prevention to one another.

Recall the proposal by Paul (1998), introduced in section 4 above, that events that change the time of occurrence of other events have a special causal status. Lewis (2000) uses this as a point of departure for his own theory of causation as influence. Talking of Paul's proposal, he writes:

> I favor this further emendation…But I think we should go further still. What's so special about time? When we thought that without the actual causes of his death, Ned Kelly would have died a different death, we were thinking not just that he would have died at a different time, but also that he would have died in a different manner.

But there *is* something special about time. In introducing his concept of *alterations* of an event, Lewis ignores a crucial distinction. Alterations that differ with respect to their time of occurrence typically stand to each other in *causal* relations, while those that do not, stand to each other in some other kind of relation of mutual exclusion.[16] If we ignore this distinction, we will simply be unable to accurately represent the causal structure of the world.

The question we have ultimately answered – how to model situations like the one described by Bennett (1987) – is not literally identical to any of the ones with which we began: whether the time of its occurrence is part of the essence of an event; whether hasteners and delayers are always causes. But the answer, I think, at least partially vindicates the positions of Lombard (1986) and Paul (1998). There *is* something special about time.[17]

---

[16] This is not *necessarily* true. A world very different from ours might have simultaneous causation, and hence simultaneous prevention as well. In such a world, we might be able to sensibly say that the gas knob's being set to high prevents it from being set to low at the very same time. And I think that a case can be made that distant correlation phenomena of the sort one finds in quantum mechanics can involve non-causal exclusion relations between spatiotemporally separated events.

[17] For comments and discussion, I'd like to thank David Baker, Ronald Giere, Ned Hall, Benj Hellie, Boris Kment, Laurie Paul, Peter van Inwagen, Jessica Wilson, participants at

the workshop on Scientific Philosophy at the University of Tilburg, and participants in the Carolina Metaphysics Workshop.

References

Bennett, J. (1987) “Event Causation: the Counterfactual Analysis,” *Philosophical Perspectives* 1.

Carnap, R. (1950a) *Logical Foundations of Probability.* Chicago: University of Chicago Press.

Carnap, R. (1950b) “Empiricism, Semantics, and Ontology,” *Revue Internationale de Philosophie* 4: 20 – 40.

Carnap, R. (1963) “Replies and Systematic Expositions,” in Schilpp (1963), pp. 859 – 1013.

Davidson, D. (1967a) “The Logical Form of Action Sentences,” in N. Rescher, ed., *The Logic of Decision and Action* (Pittsburgh: University of Pittsburgh Press).

Davidson, D. (1967b) “Causal Relations,” *Journal of Philosophy* 64.

Friedman, M. (2001) *Dynamics of Reason*. Stanford CA: CSLI Press.

Halpern, J. and C. Hitchcock (2010) “Actual Causation and the Art of Modeling,” in R. Dechter, H. Geffner, and J. Halpern (eds.), *Heuristics, Probability, and Causality: a Tribute to Judea Pearl* (London: College Publications), pp. 383 – 406.

Halpern, J., and J. Pearl (2005) “Causes and Explanations: A Structural-model Approach — Part I: Causes,” *British Journal for the Philosophy of Science* 56: 843 - 887.

Hitchcock (1993) “A Generalized Probabilistic Theory of Causal Relevance,” *Synthese* **97**: 335 - 64.

Hitchcock, C. (2001) “The Intransitivity of Causation Revealed in Equations and Graphs,” *Journal of Philosophy* 98: 273-299.

Hitchcock, C. (2004) “Routes, Processes, and Chance Lowering Causes,” in P. Dowe and P. Noordhof (eds.) *Cause and Chance* (London and New York: Routledge), pp. 138 – 151.

Jackson, F. (1998) *From Metaphysics to Ethics: A Defense of Conceptual Analysis. Oxford: Oxford University Press.*

Kant, I. (1781/1787) *A Critique of Pure Reason.*

Kim, J. (1973a) “Causation, Nomic Subsumption, and the Concept of an Event,” *Journal of Philosophy* 70: 217 – 36.

Kim, J. (1973b) “Causes and Counterfactuals,” *Journal of Philosophy* 70: 570 – 572.

Lewis, D. (1973) “Causation,” *Journal of Philosophy* 70: 556 – 567. Reprinted in Lewis (1986b), pp. 159 - 172.

Lewis, D. (1976) “The Paradoxes of Time Travel,” *American Philosophical Quarterly* 13: 145 – 52.

Lewis, D. (1979) “Counterfactual Dependence and Time's Arrow,” *Noûs* 13: 455-476. Reprinted in Lewis (1986b), pp. 32 - 52.

Lewis, D. (1986) “Events,” in Lewis (1986b), pp. 241 – 269.

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

Lewis, D. (2000) “Causation as Influence,” *Journal of Philosophy* 97: 182 - 197.

Lombard, L. (1986) *Events: A Metaphysics Study*. London: Routledge.

Manning, G. (Forthcoming) “When the Mind Became Un-natural: de la Forge and Psychology in the Cartesian Aftermath” in P. Bakker and C. Leijenhorst, eds., *Psychology and the Disciplines* (Leiden: Brill)

Paul, L. (1998) "Keeping Track of the Time: Emending the Counterfactual Analysis of Causation," *Analysis* 58: 191 – 8.

Paul, L. (This volume) "The Handmaiden's Tale: Ontological Methodology"

Pearl, J. (2000/2009). *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press (Second Edition, 2009).

Popper, K. (1959) *The Logic of Scientific Discovery*. London: Routledge.

Reichenbach, H. (1949) *The Theory of Probability*. Cambridge: Cambridge University Press.

Schulte, O. (1999) "Means-Ends Epistemology," *British Journal for Philosophy of Science* 50:1 – 31.

Spirtes, P., Glymour, C., and R. Scheines (2000) *Causation, Prediction and Search*, Second Edition. Cambridge, MA: M.I.T. Press.

Schilpp, P., ed., (1963) *The Philosophy of Rudolf Carnap*. La Salle, IL: Open Court Press.

Strawson, P. (1963) "Carnap's Views on Constructed Systems versus Natural Languages in Analytic Philosophy," in Schilpp (1963), pp. 503 – 518.

Woodward, J. (2003) *Making Things Happen.* Oxford: Oxford University Press.

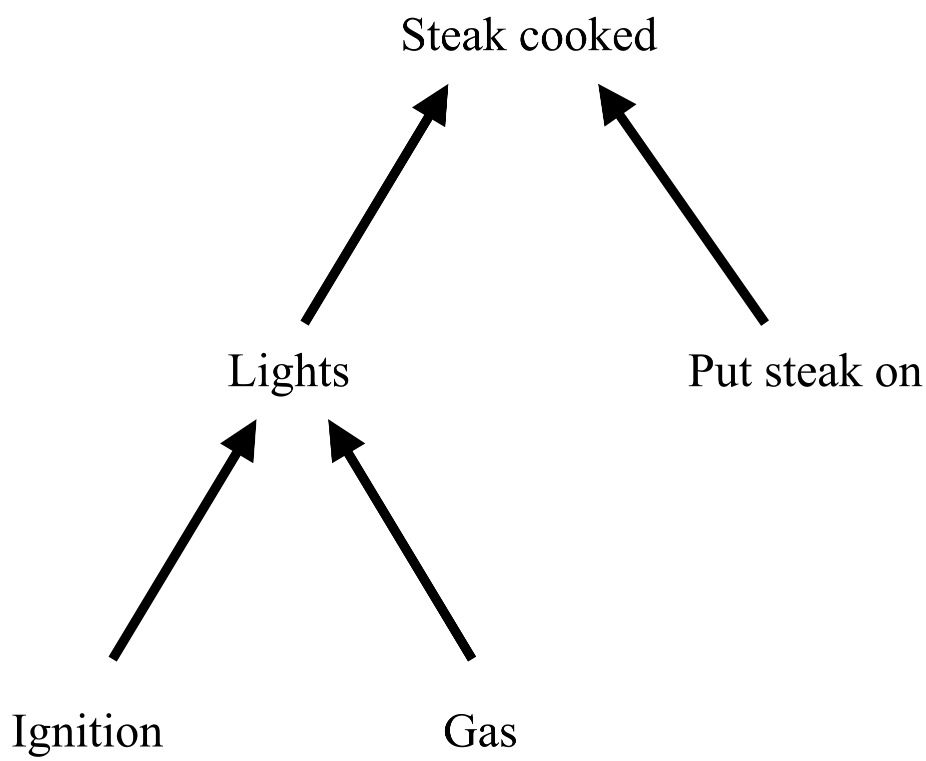

Figure 1

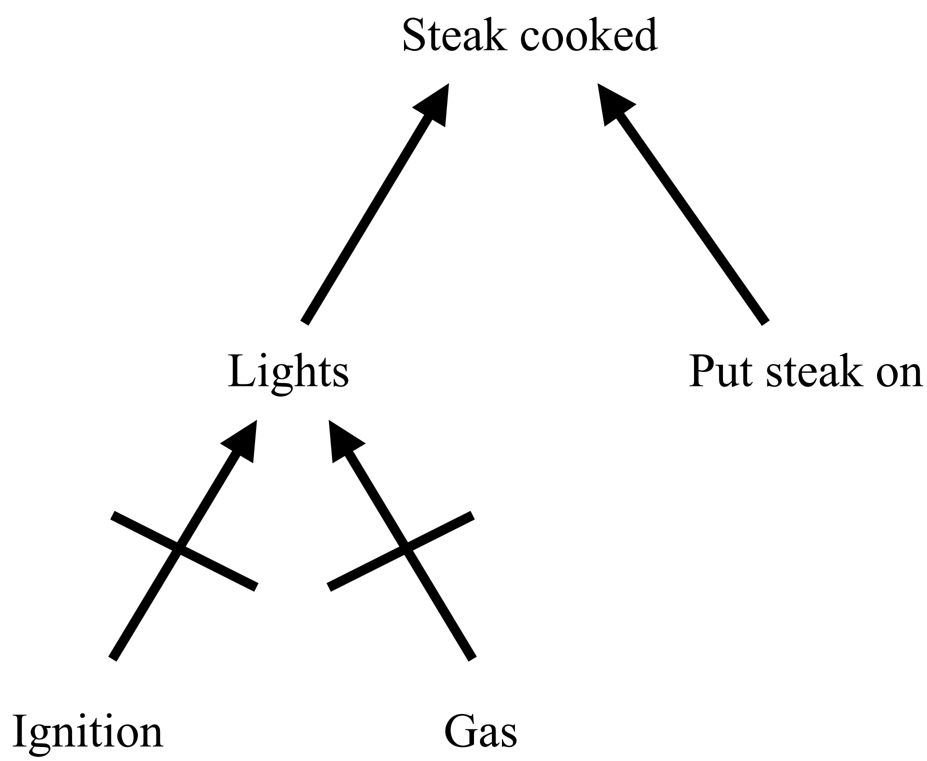

Figure 2

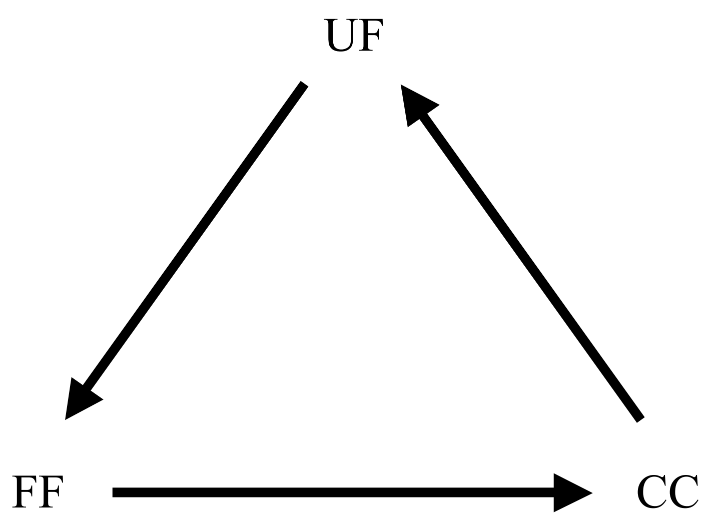

Figure 3

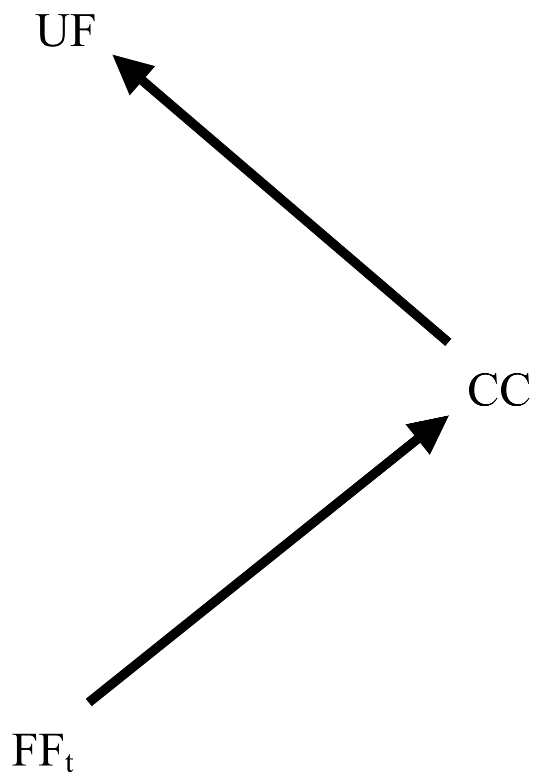

Figure 4

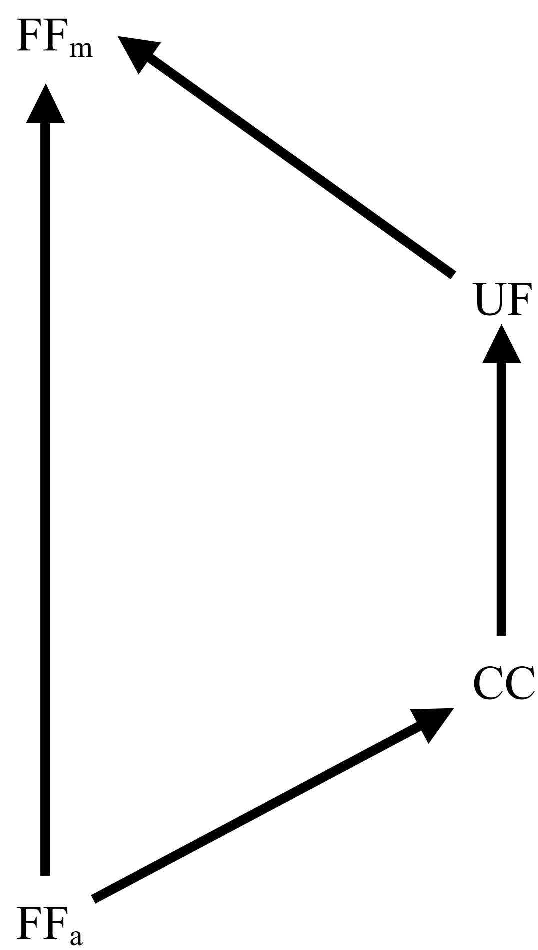

Figure 5