EMPIRICAL ANALYSES OF CAUSATION

DOUGLAS KUTACH

Abstract. Conceptual analyses can be subdivided into two classes, good and evil. Empirical analysis is the good kind, routinely practiced in the sciences. Orthodox analysis is the malevolent version that plagues philosophical discourse. In this paper, I will clarify the difference between them, provide some reasons to prefer good over evil, and illustrate their consequences for the metaphysics of causation. By conducting an empirical analysis of causation rather than an orthodox analysis, one can segregate the genuine metaphysical problems that need to be addressed from the many pseudo-problems that have long dogged traditional accounts of causation.

1. Introduction

Imagine a psychologist who has formulated a theory of how people understand various interactions among physical stuff, i.e, an account of the implicit folk theory of physics. His model incorporates parameters for characterizing contingent conditions like a value for how dense an object is represented as being or an implicit estimate of how quickly a certain object will return to rest after being set in motion. It includes hypotheses about variances among people and performance limitations that affect how people's understanding is applied in practice. Suppose all the psychologist's work is methodologically unimpeachable and that the model is stunningly successful given the criteria psychologists use for evaluating theories, e.g., making precise and accurate predictions of people's responses to questions about physics and predictions about how they will behave when confronted with practical problems that test what they know about physics.

Now imagine the response our psychologist would receive if he suggested to the physics department that his psychological theory ought to be adopted as a constraint on their theories of force and energy and so forth. The physicist according to this program would be tasked with filling in the psychological theory's various parameters to arrive at a model that matches the structure of the external world. Or worse, imagine our psychologist arguing that regardless of any virtues the physicists' current theories have, they cannot concern genuine energy because in order for any theory to be pertinent to the topic of energy, its claims would need to avoid conflict with folk wisdom concerning energy, e.g., that exercise increases a person's energy. Whether a psychology of folk physics serves as good constraint on theories of real world physics is of course ultimately an empirical question, but not only do we have independent reasons to reject this program as an extremely implausible strategy for improving physics, there is no reason to believe that a successful physics *must* obey the implicit logic of folk physics (or naive opinions about the use of physics terms) on pain of not really being a theory of physics.

Metaphysicians of causation routinely practice activities analogous to this hypothetical psychologist. All too frequently, theories designed to accommodate linguistic features of natural language are pressed into service as constraints on theories about the behavior of

Date: May 24, 2010.

the external world, with similar prospects for success. Metaphysical theories of causation are standardly required to concern the external world in the sense of being applicable to astronomy, ecology, and economics while at the same time vindicating the literal truth of folk intuitions about causes. The practice primarily stems from the routine use of a crippled form of conceptual analysis. While conceptual analysis of some sort is necessary for any useful intellectual investigation, malignant versions of it exert widespread influence over standard practices, including those of scholars who nominally disavow conceptual analysis.

The presence of the bad kind of conceptual analysis is at least understandable in philosophical disciplines having little relation to science. What is striking about the kind of conceptual analyses standardly presupposed in the philosophical literature on causation, though, is that an alternative form of conceptual analysis is readily available: empirical analysis. What empirical analysis consists in, I think, has not yet been adequately articulated, as demonstrated by continuing puzzlement over its aims, e.g., (Bontley 2006). The negative part of my task here is to expose that aspect of the orthodox metaphysics of causation which should be rejected by serious investigators of causation, and the positive part is to sketch a viable alternative to the orthodox methodology.

2. A Case Study

Nothing illustrates the odd character of the orthodox methodology better than David Lewis's (1979) theory of counterfactual asymmetry, which is tied to causation by way of his counterfactual accounts of causation (Lewis 1976, 2004). In the papers on causation, Lewis offers criteria to use in conjunction with his counterfactual logic (1973) in order to take some possible world including an event of interest e, and to determine whether a given event c is a cause of e in that world. The criteria discussed in the counterfactual asymmetry paper consist of a ranked list of respects in which two possible worlds can be compared to determine which is more similar to the actual world.

Ignoring whether the theory is successful on its own terms, it ought to strike anyone as a peculiar way to investigate causation. Lewis's grand argumentative structure involves first developing a logic of counterfactual conditionals by comparing the consequences of various axioms with suitably informed folk judgments about sample sentences and inferences; and second, examining the logic to find that it has a semantics involving a relation of comparative similarity among possible worlds. This is standard procedure in regimenting the logic and semantics of natural language and is by itself unproblematic. But in a curious third step, Lewis uses the semantic structure to concoct a system that generates truth values for the counterfactual conditionals relevant to the evaluation of causal claims. These counterfactuals are then fed into an account of causation to render judgments about instances of causation among ordinary events.

The success of the account, orthodox experts think, is to be judged by at least two crucial criteria. The first concerns how well it reproduces moderately refined folk judgments about specific cases of causation and general truths about causation, e.g., that effects never precede their causes (except perhaps in unusual environments). For every prominent account of causation x, the causation literature is replete with attacks on theory x of the form, "In such and such scenario, x's identification of the causes of e conflicts with a common sense identification of the causes." It is possible for a successful account to oppose wildly popular judgments, but such disagreements are broadly presumed to count against the theory, not the intuitions. Accepting the theory in the face of clear counterexamples is generally considered a course of action to be taken only grudgingly as a last resort.

The second concerns how well the account provides a non-trivial systematization of claims about causation. A metaphysical theory of causation is not supposed to be an encyclopedia of facts about causation nor a list of psychological heuristics that guide our identification of causes. Its purpose is to tell us about causation itself, especially what all instances of causation have in common. It should not merely fit the common sense intuitions in the way one draws a best fit curve through a graph of data points but should elucidate principles that connect causation to related concepts like laws, events, chance, time, as well as explain the reasonableness of our central beliefs concerning causation.

It is unclear what purpose is served by this methodology, especially its dismissal of accounts of causation that do not meet both standards. Some philosophers construe Lewis's project as an early *scientific* theory of causal influence.[1] So understood, the project audaciously proposes to uncover a theoretical structure that systematizes natural language, specifically our implicit counterfactual logic, and then to use that same structure to explain how causal influence in the external world operates. As illustrated by the psychologist who wants to impose his theory as a constraint on physics, such maneuvers are fantastically implausible as guiding principles for a scientific program. Of course, the mere fact that a logical structure systematizes natural language does not by itself mean it is likely to be useless for science; propositional logic is handy for understanding fragments of ordinary language and for science as well. The implausibility comes from the counterfactual logic's ability to account for idiosyncratic features of natural language. That Lewis's own logic of counterfactuals is meant to explain features of human language that are unmotivated from a physical perspective is evident from, among other things, its use of the centering axiom (which results in a significant difference between the truth conditions when the antecedent is true versus when it is false), its treatment of the conditional as a variably-strict modal operator (which conflicts with a natural treatment of chancy influence), and its treatment of negation.[2] It is even more audacious to insist that a scientific theory of causal influence is defective if its model for evaluating "what would have happened had things been otherwise" conflicts with the implicit logic of ordinary language counterfactuals.

An alternative interpretation of Lewis's counterfactual asymmetry paper tells us that it is an attempt to explicate the implicit psychology grounding our judgments of causal influence. The counterfactual logic is a constraint on theories of causation, the story would go, because one is hypothesizing that some cognitive module plays a role in our use of counterfactual language and also for causal reasoning. But, as Paul Horwich (1993) notes, Lewis's specific system is psychologically implausible because it employs facts about the amount of time two possible worlds are perfectly alike in their instantiation of properties and the spatial extent of miracles in these worlds (relative to laws of the actual world). Furthermore, construing Lewis's theory as merely teasing out an implicit psychology does not make sense of the fact that it is offered as an explanation of the causal asymmetry. One of Lewis's motivations for the theory is to avoid positing a fundamental temporal direction grounding the asymmetries of counterfactuals, influence, and causation, and instead to derive these asymmetries from facts about the contingent layout of historical fact. This makes sense if one is thinking about the account as part of a metaphysical or scientific project but not if it is just psychology. Given the psychological salience of temporal asymmetry, it

[1] I have in mind specifically David Albert's address at the 2002 Philosophy of Science Association meeting but such an interpretation also seems to be implicitly assumed by scientifically sophisticated criticisms by Elga (2001), and other similar literature that attacks Lewis's theory with technical physics.

[2] I discuss these issues elsewhere, but one can find criticisms along these lines in (Pruss 2003; Gunderson 2004; Hawthorne 2005).

is far more plausible that people simply judge causation and influence to be directed towards the future rather than continually deriving the causal direction from an estimation of durations of perfect match and sizes of miracles.

Lewis's theory is implausible as a theory of the science of causation and as a theory of the psychology of causation, but a third interpretation is that it is deliberately somewhere in between—a hybrid combining folk judgments of causation with what we know from science. This project, often known as 'conceptual analysis,' has been recently defended in general, e.g., (Chalmers 1996; Jackson 1998, Chalmers and Jackson 2001), and specifically with regard to causation, e.g., (Lewis 2004; Collins, Hall, and Paul 2004). Traditional versions of conceptual analysis were committed to the project of finding explicit definitions of folk concepts, e.g., (Ducasse 1926), but contemporary defenders allow and usually advise regimentation and improvement of the concepts. I will call the contemporary version of conceptual analysis, 'orthodox analysis' because it is de rigueur in turn of the millennium philosophical circles.

It ought to be uncontroversial that some kind of conceptual analysis is a necessary component of any intellectual investigation, for without it, we would have no way to connect our theoretical terms to the folk terms they are intended to improve and the phenomena they are intended to describe. One needs to have a requirement in one's standards for theoretical adequacy so that, for example, a successful theory of planetary motion is not passed off as a successful theory of causation merely by attaching the label 'cause' to what we ordinarily think of as an orbit. But there are better and worse versions of conceptual analysis, and despite the improvements orthodox analysis provides to old-fashioned conceptual analysis, the resulting methodology is still defective.

3. Orthodox Analysis

Food is important to our survival and flourishing, so we ought to know what things are food and what are not. This requires us to know the rules governing the extension of 'food.' In response to this challenge, a food scientist conducts a vast investigation into nutrition and finds out every detail concerning how various ingestible materials are related to creature health. The facts she uncovers, according to this fantasy, exhaustively identify every relation among every relevant variable food scientists care about. Upon completion of the scientific work, she says, "Everything we wanted to know about food is subsumed by the nutrition relations I have found. Food is basically that which is nutritious."

The orthodox analyst responds, "Your analysis is either false or off topic. The goal was to find out what food is, and your theory of nutrition only identifies the extension of your newfangled concept 'nutrient.' For your theory to be of use in understanding food, you must solve the location problem (Jackson 1998), isolating (from the morass of nutrition relations your science has identified) some structure that well enough matches folk platitudes about food. Nutrition science alone does not show which substances count as food. It must be supplemented with a conceptual analysis that provides linkage between 'nutrient' and our native food concept. To conduct such an analysis, one must compare what the theory says is nutritious with uncontroversial examples of food, as evinced from native informants or, even more efficiently, a few smart colleagues. When anyone examines your concept 'nutrient,' we find it does not match well with food. Your theory counts as nutrients tea, iron crowbars, human brains and irradiated feces, but we know a priori that tea is a beverage, a crowbar is inedible, etc. So, cheers for your theory of nutrition, but there still exists the very much unresolved project of understanding what food is."

In defending the utility of orthodox analysis, philosophers have proposed amendments (or perhaps clarifications) to make conceptual analysis seem less like a parody. Jackson, for example, correctly emphasizes the following points:

(1) One does not need to accommodate all the naive platitudes associated with the concept. One can dismiss some as mistakes or tangential aspects of the concept. Furthermore, if the full set of platitudes does not cohere, one can (presumably as a last resort) abandon enough platitudes to achieve coherence.
(2) One does not need the sought after a priori connections among concepts to take the form of an explicit definition. The conceptual analysis can handle cluster concepts by permitting somewhat hand-waving connections among concepts that draw on people's native pattern-recognition ability.

I agree that these two principles should be adopted for conceptual analysis, but they pose a problem for orthodox analysis because they appear to weaken the standards for adequate analysis so much that it becomes indistinguishable from the kind of conceptual analysis common in science, where little attention is paid to disagreements with folk opinion. An examination of the actual practices of orthodox analysts reveals that their intended form of analysis is more restrictive by insisting on the importance of rendering folk platitudes literally true. Collins, Hall, and Paul (2004, p. 31) attempt to ward off an overly permissive form of conceptual analysis by stating "although the account can selectively diverge from these intuitions, provided there are principled reasons for doing so, it should not diverge from them wholesale." Depending on one's standards for 'principled' and 'wholesale divergence,' orthodox analysis could still be interpreted as allowing the kind of revisionary conceptual precisifications common in science, but I take their claim to be an attempt to disallow such a weak construal of the standards of adequacy. What the orthodoxy must defend—in order to be distinguishable from the empirically-oriented version of conceptual analysis I will soon clarify—is its practice of imposing a heavy burden of explanation on any proponent to account for why we should reject obvious truths that conflict with her proposal.

According to official doctrine, the orthodoxy insists in general that folk intuitions and platitudes about X be taken as touchstones for judging the adequacy of analyses of X in the sense that the analysis must make them come out as strict truths and not as strictly false but entirely reasonable simplifications of reality. But in practice, the doctrine is applied in a biased fashion to accord with popular opinion. In cases like the relation between 'nutrient' and 'food,' where a strict implementation of the orthodox methodology would reveal itself as preposterous, the common sense intuitions are brushed aside as pedantic niceties. In practice, that is, orthodox analysts accept the food scientist's claims about nutrients as informing us about food despite the lack of explicit explanations of all the discrepancies between 'nutrient' and 'food.'

However, in cases where orthodox analysis serves as a useful shield for the analyst's prejudices or the status quo, the folk intuitions are held up as definitive standards, obvious truths to be abandoned only as a last resort under the force of weighty reasons. Specifically with regard to causation, the orthodox analyst is insistent that central folk intuitions be strictly respected. Every orthodox analyst demands that an adequate account of causation must respect the principles that (1) events do not cause themselves, (2) effects do not cause their causes, and (3) preempted would-be causes are not genuine causes.[3] Contradicting any of these obvious truths about causation counts as grounds for dismissal,

[3]Preempted events are by definition not causes of that which they were going to cause. What I mean is the non-trivial claim, "Those events people standardly identify as having been preempted from causing E are

unless a convincing explanation is provided for why the violated principle should be abandoned. For example, a symmetric theory of causation holds that anytime c causes e, e also causes c. One counterexample that is taken seriously by the orthodox theorist as sufficient grounds for rejecting the symmetric theory is the common sense observation that an ordinary instance of thunder is not a cause of the previous lightning. Attempts to brush off the alleged counterexample by claiming that our disinclination to identify it as a genuine cause is just a result of pragmatic factors—e.g., that we often conflate the more practically useful future-directed causation with causation simplicter—are not taken seriously without extensive explanation.

But what marks the difference between discrepancies that are so easily explained away that they hardly demand explicit discussion from discrepancies where a heavy burden is placed on the proponent? The de facto standard is that an explanatory burden exists if and only if most experts find the account's claim counterintuitive. But this rule is ipso facto incapable of distinguishing misguided intuitions shared by the bulk of the expert community from intuitions that are essential for the analysis to be correct and on topic. The crux of the problem with orthodox analysis is simply this: An orthodox analysis of X has no *principled* means for distinguishing between (1) platitudes one must accommodate because to dispense with them would guarantee either that the account is false or that it is not an analysis of X, and (2) platitudes that are misguided but strongly believed as strictly true because we humans are simple minded, easily indoctrinated, or genetically predisposed. This deficiency prevents a successful analysis whenever people share strongly held intuitions, some of which are crucial for isolating the subject of discussion and others of which are bogus beliefs due to widespread ignorance or failure to recognize the ridiculousness of prevailing doctrines, etc. In those cases, any attempt at orthodox analysis will necessarily be crippled by its obligation to vindicate the literal truth of the bogus intuitions. Once accepted as among the platitudes that concern the meaning of X, an orthodox analysis of X has no mechanism to expunge the cognitive dreck without facing the damning charge of having changed topic or having claimed something patently false. At best, an orthodox analysis can grudgingly come to accept an imperfect fit with the folk intuitions once it has become clear enough that no perfect analysis is forthcoming.

Orthodox analysis thus provides no *structured* means of escaping a conceptual trap, defined as any situation where our native concept of X includes platitudes that are in fact strictly false (or are strictly false according to some ideal account of X that we would all recognize as the best account if we were suitably informed). The orthodoxy's only means for avoiding conceptual traps is to permit some platitudes to be abandoned if there are "principled arguments for doing so." But because typically there are multiple ways to imperfectly vindicate the full set of platitudes, and orthodox analysis provides no guidance beyond gut feeling for how to prefer one platitude over another or how to balance degree of systematization against degree of platitude fit, the acceptance of orthodox analysis as the way to conduct philosophical disputes often results in fruitless squabbles over whose imperfect analysis should count as the unique best analysis.[4] By contrast, empirical analysis provides a methodology for escaping conceptual traps.

---

not genuine causes of E." Also, the orthodox analyst is free to allow that the asymmetry of causation might be violated in exceptional circumstances.

[4]See (Hitchcock 2003) for a list of such pseudo-debates in the causation literature.

4. Empirical Analysis

A small fraction of the literature on causation is directed towards a somewhat different goal from that of orthodox analysis. Though the general idea has been presented under various names, I will follow Phil Dowe by labeling the alternative approach the 'empirical analysis' of causation. The first chapter of Phil Dowe's (2000) *Physical Causation* marks an advance in clarifying this different kind of investigation of causation. In this section I will distinguish my version of empirical analysis, using Dowe's as a reference point. My differences with Dowe should be read not so much as criticism of his project but as an attempt to push his key idea much further to make a cleaner break from orthodox analysis.

Dowe characterizes the empirical analysis of causation in several ways. Its task is "to discover what causation is in the objective world." (p. 1) It "aims to map the objective world, not our concepts." (p. 3) Dowe argues that it is a mistake to demand empirical analyses account for all the ways we talk about causation, or to demand that empirical analysis "hold good for all logically possible worlds." (p. 6) Thus, it is in the business of discovering "what causation is as a contingent fact." (p. 4) Dowe contrasts it with the old-fashioned conceptual analysis espoused by Ducasse (1926).

I agree with Dowe as far as he goes, but he does not go nearly far enough to distinguish what makes empirical analysis a better way to conduct conceptual analysis. The key question Dowe does not answer is, "What distinguishes those intuitions the analysis needs to accommodate from those that it can set aside?" Like orthodox analysis, Dowe's explicit characterization of empirical analysis provides no guidance, and his own empirical analysis of causation, his conserved quantity account, does not provide many clues as to what the answer would be. On some occasions (p. 110), he emphasizes that his account does not need to accommodate all folk intuitions about causation, but on other occasions, he appeals to everyday causal talk to criticize other accounts (pp. 24, 40) and to motivate significant extensions to his own theory (p. 148). What is missing is a scheme to unpack the operational meaning of "what causation is."

Here is my attempt to characterize empirical analysis.

> An empirical analysis of X is a conceptual structure designed to optimize explanations of whatever empirical phenomena make X a concept worth having.

There is obviously a lot of vagueness in this definition and a threat of vacuity, but it is not fruitful to pretend that this statement or some more precise version of it should count as a set of informative necessary and sufficient conditions that will stake out a clear boundary between empirical analysis and orthodox analysis. Instead, one acquires the thrust of empirical analysis by abstracting from the kind of conceptual analysis done in exemplary sciences. Empirical analysis is a cluster concept, best identified as any conceptual analysis that comes close enough to paradigmatic examples of the kinds of analysis done in science. In addition to the already mentioned example from food science, I will discuss two exemplars that illustrate different features of empirical analysis. Many other sciences exhibit excellent conceptual design, but two examples should be enough to convey the basic idea. Then, we can make the methodology a bit more precise by formulating an algorithm for how to hack through any bogus platitudes that hold us in a conceptual trap.

Conducting an empirical analysis is not the investigation of some concept that we take as a preexisting object of inquiry but rather as a task of conceptual engineering. Classical physics, for example, distinguishes three kinds of mass:

- Inertial mass is a body's degree of resistance to external forces.

- Gravitational source mass quantifies how strong of a gravitational field that body produces.
- Gravitational coupling mass quantifies how much that body is affected by the gravitational field at its location.

Because all three aspects can be represented without loss of content by the same variable, classical physics treats them all as aspects of a single mass property. In the standard interpretation of general relativity (GR), the three notions come apart. Inertial mass retains its classical role, but the coupling mass drops out of the theory entirely because gravity is no longer really a force, and the gravitational source mass is replaced with a ten-component tensor. If one were conducting an orthodox analysis (working under the fiction that GR represents the final scientific story about mass) the question to be asked, according to the defender of orthodox analysis, would be, "Which structure in GR best plays the naive mass role?" The physicist rightly does not much care about such a question. The important work was getting the theory of gravity right and clarifying what structures are needed to play each role. Of course, some conceptual relation must exist between the various mass-like theoretical concepts and what we intuitively take to be masses, but an acceptable connection comes from the totality of GR's concepts through the explanations it provides. The reason for having our naive mass concept is that it allows us to think in simple terms and still get a practical theory of motion and gravity. But to the extent we are satisfied with GR's explanations, we should automatically be satisfied about the utility of the naive mass concept without needing to find which unique structure in the internals of GR best corresponds to the totality of roles we associate with mass. In an empirical analysis, it is not obligatory to solve Jackson's (1998) location problem, at least as he conceives of it.

Another illustration of empirical analysis exists in the famous debate over whether space is a substance. In classical physics, there is a clear enough distinction between two opposing camps. Anti-substantivalists believe space is merely a fiction useful for describing facts that are fundamentally about distance relations among material bodies. Substantivalists believe space exists as something in its own right, despite its being unusual (for a substance) by not acting on or reacting to other substances by way of forces. GR alters the debate by providing an empirically superior account of the physics where the distinction between matter and space (or, from here on, spacetime) becomes blurrier. Given how it handles the concept of mass, one might think empirical analysis should avoid taking sides on whether the best account of spacetime in GR is a substance: that so long as GR is a good empirical guide, whether its spacetime is substantival is a mere labeling issue. But that would be incorrect because what is philosophically important is whether the motivation in the classical debate for caring about whether space (or spacetime) is substantival is satisfied by GR. It is clear enough that the classical anti-substantivalist wanted to explain the motion of particles without taking ontologically seriously this allegedly metaphysically problematic spatial structure and to account for its conceptual utility by recovering putative facts about space from a fundamental metaphysics that is paradigmatically material. So when it turns out in GR that the spacetime structure needed to explain the motion of matter is not derivable from relations among paradigmatically material entities, it arguably posits a structure that is ontologically more than just an aspect of matter. Thus, whether spacetime is a substance is not a verbal dispute. Rather, GR vindicates substantivalism because the scientific motivation for being a substantivalist about space is that it takes more than just spatial relations among paradigmatically material things to account for motion. There remains some controversy among experts about whether this is the correct lesson to draw from GR, but my purpose here is to illustrate how empirical analysis is not merely a

disguised form of instrumentalism. There are genuine debates about what structures are to be taken metaphysically seriously, debates that are to be settled where possible by careful examination of which metaphysical system best accounts for the empirical phenomena.

The method suggested by these examples is that we begin an empirical analysis of *X* by assembling all our common sense intuitions about *X* and platitudes about the constitutive roles associated with *X*. Then we try to figure out which empirical facts make this collection useful, preferably by formulating experiments that characterize the core empirical facts. Then, we seek scientific explanations for these experiments, optimizing our concepts to improve these explanations. The advocated methodology does not require that we find a unique correct empirical fact that corresponds to the concept. One should just examine what seem like prima facie interesting empirical questions that motivate us in believing something roughly resembling the initial collection of platitudes. Conducting the analysis might motivate reconsideration of what is important or a revision in exactly what the core issue is, as happened in the change from caring about whether space is a substance to whether spacetime is a substance. Also, because there is no way to think about empirical facts without conceiving of these facts in some way, there is always some extent to which intuitions and naive beliefs about reality inescapably affect one's conceptual analysis. Finally, it is important that empirical analysis be understood in a way that does not require a sharp distinction between what is empirical and what is conceptual because such a distinction cannot be made precise for the intended notion of 'empirical' and the success of science in general does not depend on it. For example, we might think, "Does spacetime exist?" is an empirical question in the sense that we can empirically assess the relative prospects of spacetime theories versus competitors that posit no spacetime. But we could also think of the existence of spacetime as not being an empirical issue for a variety of reasons. Science might result in two equally acceptable, ideally adequate fundamental theories, one of which treats spacetime structure as a fiction and the other of which treats it ontologically seriously. If so, our inability to experimentally check whether spacetime exists casts no serious doubt on the quality of the explanations the two theories provide, and thus no doubt on the utility of the concepts honed to improve these explanations.

Almost always, the initial platitudes concerning *X* will cluster into two core groups: those associated with *X* itself (as something out there in the real world) and those associated with our psychology of *X*. For example, our initial platitudes about food might include that it is the kind of thing that (1) we require for survival, (2) share with guests, (3) is not gaseous, and (4) typically provokes a "Yes" response by English speakers who are asked, "Is this food?" When we think of why we care about food, it is obvious enough that its role in our survival and proper physiological development is of primary importance and its role in facilitating social bonds is parasitic on its utility for survival. The experiments clarifying the empirical phenomena are obvious: People who eat the normal amount of paradigmatic foods survive better than people who ingest similar amounts of paradigmatic non-foods like rubber or wood. The correct skeletal explanation—that food is composed of molecules that promote survival—motivates us to use a regimented concept of food, which we then optimize to fit better with facts that we did not initially include as part of the food platitudes. Oxygen molecules promote survival too, and so does iron, which might motivate us to count them as nutrients. But as we optimize 'food' towards 'nutrient,' we generate greater discrepancy with platitudes like (2), (3) and (4). For orthodox analyses, such discrepancies put pressure on us either to say a block of iron is not food or that nutrients are not (near enough) the same thing as food. For empirical analyses, we treat such platitudes as irrelevant to the explanation of what was really important about food,

and instead delegate them to a psychological study of our native food concept. Where orthodox analysis indiscriminately mixes platitudes about X and about the psychology of X, the empirical approach instructs us to segregate the platitudes into these two groups and to systematize each group separately using an empirical analysis.

5. The Empirical Analysis of Causation

So much for empirical analysis in general. What are its implications for the metaphysics of causation? The goal of inquiry into the metaphysics of causation is to find scientific explanations for whatever empirical phenomena make causation a concept worth having. An empirical analysis of causation is just the collection of concepts optimized for such explanations. To conduct an empirical analysis, one should isolate whatever phenomena vindicate our use of causal concepts, and then try to extract some characterization of those phenomena in terms of stuff whose empirical status is not controversial.

Rather than survey the full space of possibilities, it is useful for illustrating the power of empirical analysis that we just examine one commonly mentioned reason for believing in causal structure: that there exist "effective strategies," in Nancy Cartwright's (1979) phrase, for influencing the world. Creatures like us, who behave in paradigmatically agential ways, are able to manipulate events, including what we directly control and what we indirectly influence. For this to be true, there needs to be at least some regular structure in how various bits of the world are generally correlated with our actions. But in what sense is that backed by something empirical? Suppose we have a bunch of nearly identical experimental setups instantiating an agent embedded in an environment. Let S be the event type representing what is common in such setups. Half of the setups involve the agent performing an action of type A_1 and the other half A_2. The empirical import of "effective strategies" can be interpreted as the fact that there exist a vast number of event types S, E, A_1 and A_2 such that E happens more often when things start with $S + A_1$ than with $S + A_2$. There is a lot of fuzziness concerning what kinds of event types are being claimed to exist. While there is no need for great precision, enough of them need to be epistemically identifiable and expressible using human concepts, so that the types are not generally gerrymandered in a way that trivializes the existence of effective strategies.

So far, we have identified a fact that is prima facie empirical, but characterizing effective strategies in terms of agency might raise a worry that the employed notion of agency incorporates something non-empirical into our explanandum. If the invoked notion requires some epistemically inaccessible aspect of reality, we would have failed to distill the "effective strategies" idea into a satisfactory basis for empirical analysis. To ensure that agency is empirically kosher, one should show that agency can be construed in a way that is no more mysterious than any ordinary physical functionality. One argument involves demonstrating that along a continuum as one considers entities that are less and less agential, the features about agency that appear in the characterization of effective strategies degrades gracefully. Even if one wants to designate a precise boundary between agents and non-agents for the purposes of logic or semantics or ethics, one would still like it to be the case that the empirical behavior of a pair of very nearly identical entities—one just barely an agent, the other just barely a non-agent—differs only because of their material constitution, not because the evolution of the material world treats agency itself as significant. To check whether agency degrades gracefully, we can consider agents so crude that they hardly deserve to be called agents, and see what "effective strategies" means for them. For example, although volcanos do not literally formulate and execute strategies, it is still true that there is some objective structure in the world such that the action of volcanos makes

a difference in whether lava is spread around. Examine a lot of volcanos that are similar except for whether they are erupting. The empirical upshot of the volcanos' "ability" to spread lava consists in the fact that the erupting group of volcanos is followed shortly by more lava having been spread around than in the non-erupting group. This captures the essence of the "effective strategies" idea without invoking any suspicious kind of agency.

One should continue unpacking the content of any seemingly empirically dubious element, thoroughly rooting them out until one gets to a basis that is scientifically uncontroversial. This does not require settling on some specific type of empirically fundamental entity. One just uses whatever standards are scientifically acceptable, and any disputes are delegated to epistemology or the theory of perception. For this brief illustration, I will assume the empirical content of causation can be stripped down to some facts about the layout of paradigmatically material stuff, including some laws of nature.

With the basic structure in place, one can review other aspects of causation to see whether there is a corresponding collection of empirical phenomena motivating it. In the case of the asymmetry of causation, there is an obvious family of possibilities to explore, which is that while some effective strategies exist for influencing the future, apparently there are none for influencing the past. To flesh out this idea one should try to formulate an experiment that reveals the facts to be explained. For a simplified example, imagine some event-kind E and a bunch of agents randomly assigned either the goal of having E occur or the goal of having E not occur. A plausible candidate for the empirical upshot of the causal asymmetry is that (1) for many event-kinds E in the future, E occurs reliably more often for agents trying to accomplish E than for agents who are trying to avoid E; and (2) for any E located towards the past, E happens just as often when the agent's goal is E as when the agent's goal is not-E.[5] I will not argue here that this is the best way to think about the asymmetry of causation but merely point out that characterizing such experiments is crucial to a proper empirical analysis.

Empirical analysis offers us a significant advantage over orthodox analysis in understanding causal asymmetry. If we take every common sense intuition seriously as a touchstone, that requires an analysis to make literally true such parcels of wisdom as, "Effects never precede their causes," or "Present facts do not causally influence the past."[6] With empirical analysis, we do not need to assume these naive intuitions are strictly true. We only need to make sense of the phenomena that make these claims seemingly reasonable things to believe if you haven't bothered to think deeply about the issues involved. This helps to explain how a causal asymmetry among macroscopic stuff is compatible with determinism without positing a fundamental direction of causation. One could argue, e.g., (Kutach 2010), that we routinely influence the past, but because such influence is unexploitable for accomplishing goals, it is cognitively convenient and mostly harmless to think of the past as immune to influence.

For another example of how empirical analysis leads us away from the orthodoxy's demand to validate naive intuitions about test cases, we need only consider what empirical phenomenon grounds cases of preemption. Preemption occurs when there are two potential causes, C_1 and C_2, of a single effect E, and one of them, say C_2, stops C_1 from causing E. For concreteness, suppose at some initial time there is an event C_1, a slow moving rock on a trajectory that it is initially 90% likely to break a certain window. Shortly thereafter, C_2

[5]Of course, actual distributions will reveal a difference due to ordinary statistical error. My prediction is that we will only find as many correlations as our theory of statistical error predicts.

[6]Again, the orthodoxy can permit exceptions for unusual circumstances where these principles are violated, like in the presence of space-time wormholes or time travel machines.

occurs: someone throws a rock that is very fast and has only a 1% chance of breaking the window. As chance has it, the second rock succeeds in breaking the window quickly, thus making the first rock pass through the broken window without touching any glass. Folk intuition dictates that C_1 is not among the causes of the breakage, and the orthodoxy insists that any successful theory of causation make that folk intuition come out literally true.

Suppose everyone grants the following two classes of facts as among those which are empirically accessible.

(1) There are the fully detailed singular facts about the rocks, window, and the environment, i.e., the full microscopic history.
(2) There are also general facts about what the laws entail for any hypothetical setup, including any chances that the laws fix for future events.

The first is empirical in the ordinary scientific sense where every actual chunk of physics is individually epistemically accessible in principle. The second is empirical in the sense that we can learn about the laws of nature, and we can learn about what they entail for given initial conditions by running multiple trials with the same initial conditions, and inferring the chances from the outcome frequencies. Of course, there are limitations in our ability to establish desired initial conditions and to correctly infer chances from frequencies, but such limitations are routine in science and so do not threaten the kind of epistemic accessibility we need for empirical analysis.

Notice that the preemption example presumes that C_1 is a cause in the very weak sense that it instantiates something that plays a part in the overall physical development of nature towards E. It is also a probability-raiser of E because the presence of C_1 (rather than no rock at that location) makes the probability of E be roughly 91% (rather than 1%). When the orthodox metaphysician of causation says that C_1 is not a cause, he does not mean that it plays absolutely no role in E's coming about, for it exerts a gravitational influence if nothing else and it uncontroversially affects the chance of E. The orthodox analyst is claiming that C_1 is not a cause in the pertinent sense. Now the question to ask is whether there is anything empirical to C_1's alleged status as a non-cause that goes beyond the singular fact that it played a role in the development of reality towards E and general facts about chances, that broken windows are more likely when a slow but accurate throw is made than in situations that are identical except without the throw.

It is certainly an empirically testable fact that if you present people with the description of the preemption example, they will identify C_1 as not being a cause of E, so there exist empirical facts that need to be accounted for. But these are the proper subject of a psychology of causation. The metaphysically relevant issue is whether there is something in the external world that verifies the folk claim that C_1 is not a cause of E beyond just making it generally reasonable for folk to have the kind of rough and ready notion of singular causation that includes intuitions about preemption. The method of empirical analysis tells us that the way to figure out whether there is a metaphysical fact of the matter that C_1 is not a cause is to figure out whether we can explain the existence and general character of effective strategies just using the two classes of facts listed above, i.e., without using or implying additional facts about which events were "genuine" causes of E. If so, the intuition that C_1 was preempted does not correspond to anything in the metaphysics but is just a psychological artifact. This illustrates skeletally how empirical analysis provides a principled methodology for getting out of the conceptual trap. If we determine that facts about preemption play no role in explaining the empirical phenomena that give us a reason to think in causal terms, we can set aside platitudes about preemption as metaphysically irrelevant. That makes it much easier to figure out what causation is "in the objective world."

Orthodox analysis in the metaphysics of causation sets for itself the task of satisfying platitudes concerning our somewhat folksy identification of causes as well as platitudes concerning the relation of causation to time and laws and other things that are uncontroversially part of metaphysics. The method of empirical analysis suggests we should break apart the set of platitudes into those that concern our psychology of causation and those that concern the metaphysics of causation. By using two empirical analyses, one of causation and another of the psychology of causation, a more optimal conceptual design can be achieved. Because the metaphysically oriented concepts are not held captive to naive intuitions about singular causes among ordinary events, one gets a cleaner account of how determination and probability-raising explains the existence and character of effective strategies. Because the psychologically oriented concepts do not need to do any metaphysical work, they can be treated in a more hand-waving fashion, without demanding that such intuitions be ultimately coherent.

6. Conclusion

Returning to the food analogy, there is one respectable project of uncovering that which is nutritious. Another respectable project is to figure out people's psychology of food, i.e., why they categorize certain items as food and others as non-food. Whatever that story is, it almost certainly is going to involve as a first approximation that our food concept roughly tracks that which is nutritious. At a second order approximation, facts about perception, culture, the need for cognitive efficiency, and a whole bunch of other factors irrelevant to nutrition are going to come into play to explain why 'what is food' is not precisely the same as 'what is nutritious.'

Analogously, one respectable project is to find out how the external world is structured such that some events serve as good means for bringing about other events. That constitutes the metaphysics of causation. Another respectable project is to figure out people's psychology of causation, why they categorize certain happenings as causes and others as non-causes. Whatever that story is, it almost certainly will involve a first approximation that the causation concept roughly tracks whatever is responsible for the existence of effective strategies and general facts about them, e.g., that effective strategies are temporally asymmetric. At a second order approximation, facts about perception, our need to learn about causal regularities without running controlled experimental trials, cognitive efficiency, and perhaps even culture are all going to come into play to explain why 'what was the cause' is not equivalent to 'what was driving the world's temporal evolution.'

References

[1] Bontley, T., (2006). "What is an Empirical Analysis of Causation?" *Synthese* **151**, 177–200.

[2] Cartwright, N. (1979). "Causal Laws and Effective Strategies," *Noûs* **13** 419–437. Reprinted in N. Cartwright (ed.), *How the Laws of Physics Lie* (Oxford: Clarendon Press, 1983, 21–43).

[3] Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.

[4] Chalmers, D. and Jackson, F. (2001). "Conceptual Analysis and Reductive Explanation," *The Philosophical Review*, **110** (3), 315–360.

[5] Collins, J. Hall, N. and Paul L. A. (2004). *Causation and Counterfactuals*. Cambridge: MIT Press.

[6] Dowe, P. (2000). *Physical Causation*. Cambridge: Cambridge University Press.

[7] Elga, A. (2001). "Statistical Mechanics and the Asymmetry of Counterfactual Dependence," *Philosophy of Science* **68** (3) Supplement: Proceedings of the 2000 Biennial Meeting of the Philosophy of Science Association. Part I: Contributed Papers, S313–S324.

[8] Gunderson, L. B. (2004). "Outline of a New Semantics for Counterfactuals," *Pacific Philosophical Quarterly* **85**, 1–20.

[9] Hawthorne, J. (2005). "Chance and Counterfactuals," *Philosophy and Phenomenological Research* **70** (2), 396–405.
[10] Hitchcock, C. (2003). "Of Humean Bondage," *British Journal for the Philosophy of Science* **54**, 1–25.
[11] Horwich, P. (1993). "Lewis's Programme," in *Causation*, Sosa, E. and Tooley, M. (eds.), Oxford: Oxford University Press.
[12] Jackson, F. (1998). *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Oxford University Press.
[13] Kutach, D. (2010). "The Asymmetry of Influence," *Oxford Handbook on Time*. Oxford: Oxford University Press.
[14] Lewis, D. (1973). *Counterfactuals*. Oxford: Blackwell.
[15] Lewis, D. (1976). "Causation," *The Journal of Philosophy* **70**, 556–67.
[16] Lewis, D. (1979). "Counterfactual Dependence and Time's Arrow," *Noûs* **13**, 455–76, reprinted in *Philosophical Papers, Volume 2*, Oxford: Oxford University Press, 1986.
[17] Lewis, D. (2004). "Causation as Influence," in *Causation and Counterfactuals*, J. Collins, N. Hall and L. A. Paul (eds.) Cambridge, MA: MIT Press.
[18] Pruss, A. (2003). "David Lewis's Counterfactual Arrow of Time," *Noûs* **37** (4), 606–37.