Counterfactual Support: Why Care?

Michael Strevens

Draft of October 2013

Abstract

It seems very important to us whether or not a generalization offers counterfactual support—but why? Surely what happens in other possible worlds can neither help nor hurt us? This paper explores the question whether counterfactual support does, nevertheless, have some practical value. (The question of theoretical value will be addressed but then put aside.) The following thesis is proposed: the counterfactual-supporting generalizations are those for which there exists a compact and under normal circumstances knowable basis determining the fine-grained pattern of actual variation between the properties associated by the generalization (e.g., for the generalization *Fs tend to be G*, the exact circumstances under which any particular *F* is *G*); further, the better we understand the basis and the scope of the support offered, the greater our knowledge of fine-grained variation. We care about counterfactual support because we care about actual fine-grained variation.

1. The Significance of Counterfactual Support

Let a raven's *year-point* be the precise spatial position of its center of mass exactly one year after its hatching.[1] Let Y be the set of all year-points of all actual ravens, past and future. Then the universal generalization *All ravens have a yearpoint in* Y is true; to put it another way, there is an exceptionless connection or correlation between the properties of ravenhood and having a Y-yearpoint.

There is also a nearly exceptionless connection between the properties of ravenhood and blackness. Of the two regularities, however, the one with exceptions—*All ravens are black*—interests us more. Even were we somehow to learn about the connection between ravenhood and having a Y-yearpoint, we would consider it no more worthy of scientific regard than the precise number of grains of sand right now on Ninety Mile Beach in New Zealand, or other compendious yet somehow weightless facts. Why? Because, a venerable and I think correct story goes, the connection between ravenhood and blackness is robust, whereas the connection between ravenhood and Y-yearpoint*s* is not, or to put it another way, *Ravens are black* offers a considerable degree of counterfactual support whereas *Ravens have* Y*-yearpoints* offers none.[2]

Some writers have explained the distinction by proposing that counterfactual support is diagnostic of "lawlikeness". In my view—not to be defended here—it is counterfactual support that really matters to science. You might with some reason doubt that *All ravens are black* is a law, perhaps because, unlike Newton's laws, it lacks broad systematizing power. But you would still have to grant it a scientific importance that the yearpoint generalization

1. The vagueness of the notions of ravenhood and hatching puts a lower limit on a yearpoint's possible precision; do not let this bother you.

2. If you do not think *All ravens are black* is very significant, then try a different pair, such as *All protons have a charge of* 1.602×10^{-19} C and *All protons have Y2K-states in* Y, where a Y2K-state is the quantum state of a proton in the first moment of the year 2000 UTC, and Y is the set of the Y2K-states of all actual protons. Or do something similar with quantum fields.

entirely lacks, because it offers counterfactual support by way of conditionals of the following form:

> If these two ravens (pointing to healthy male and female ravens) were to have mated, their offspring would have been black;
>
> If this raven had been raised on a diet of bacon double cheeseburgers (pointing to a healthy raven) it would still have been black;

and so on.[3] Counterfactual support, then—and the more the better—means an awful lot to science and, I would add, an awful lot to laypeople too.

Maybe you do not agree with some of this. Maybe you do not care very much about ravens. Not to worry. My main point is that to most scientists, to most philosophers, and to most ordinary people, that a generalization offers counterfactual support is tremendously significant.

There is a philosophical question and a sociological question that you might ask about this wholehearted regard for unrealized what-ifs. The philosophical question is whether science ought to care about counterfactual support. The sociological question is why scientists and other ordinary people do care about counterfactual support. I want to approach both of these questions by asking a third, related question: what is the practical value, if any, of knowledge of a generalization's counterfactual support? (I take such knowledge to be not only knowledge *that* a generalization offers counterfactual support, but also knowledge of when and how it offers such support.)

Of course, a negative answer to this question—an argument that knowledge of counterfactual support is without practical value—would settle neither the philosophical nor the sociological question, but what I aim to provide is a

3. If this raven had been bleached, it would not have been black, but we regard such cases as outside the scope of the raven blackness generalization; thus, a bleached raven is not a counterexample to *All ravens are black*, perhaps because of an implicit rider restricting the generalization's domain to naturally colored ravens (Strevens 2012). There are, however, some counterfactuals that the raven generalization really does fail to support (see section 3); the generalization therefore does not offer blanket support. It does, however, offer enough support to be scientifically interesting.

positive answer, an explanation of how counterfactual knowledge is useful for prediction and control. In the present piece I will stop there, leaving it to you to assess the significance of counterfactual knowledge's practical utility for the questions of why we do and whether we ought to invest such significance in a generalization's provision of counterfactual support.

Some circumscription: you will not find, in what follows, an explanation of the practical value of modal thinking, or even of the value of counterfactual thinking, that is, thinking using counterfactual conditionals. It is the significance of generalizations' provision of counterfactual support alone that is my explanatory object. I will, nevertheless, have something illuminating to say about the way that the truth of counterfactual conditionals is determined, even if I do not explain why we find them useful—which is to say that, although I will not explain why we think counterfactually, I will try to explain to some extent why we think counterfactually in the particular way that we do.

I have given only a very loose characterization of what it is for a generalization to provide some degree of counterfactual support. Much more will be said on this question in section 3. Before I dive into these details, however, let me consider some arguments for and against the importance of support.

2. Enter the Actualist

2.1 The Actualist Challenge Consider the following line of thought, beamed to this paper directly from mid-twentieth century logical empiricism, pressing the adequacy of actualism in all scientific endeavors. (By *actualism* I mean the dismissal of all modal talk as irrelevant, non-existent, or incoherent. The word also names a view about the metaphysics of possible worlds or possibilities held by many philosophers who value modal talk highly; my actualists are against the modal in all its forms, except perhaps conceptual necessity.)

Assume that every generalization can be separated into an "actual" and a "modal" part, with the actual part stating the consequences of the generalization in the actual world and the modal part stating the rest. (You might distinguish a third part containing the generalization's logical or mathematical content; this part stays out of the debate altogether.) For example, *All ravens are black* can be divided into a part stating that all actual ravens in actual circumstances are black and a part stating that certain counterfactual ravens (such as the fruit of counterfactual matings) and certain actual ravens in counterfactual circumstances are black. Then, the actualist urges, for all practical and theoretical purposes the actual component of a generalization is what matters. We ought therefore to abandon any sentimental attachment we have to the modal part. Indeed, it is a mystery how such an attachment could ever have developed, given that the big cognitive engineer in the sky—evolution by natural selection—is of a rather utilitarian cast of mind.

The actualist argument operates, in its natural form, on a large scale, calling into question the importance of counterfactual thought, or modal thought more generally, wherever it is found. In this paper, it can be given a narrower focus. Even given that we have good reason to think modally, even given that we have reason to think counterfactually—that is, to pose questions of the form "What if things had been different?" and to answer them using counterfactual conditionals—you can still ask: why do we consider the modal part of a generalization to be important, scientifically or practically? And I do ask it.

Let me consider, rather too briefly, several answers, sketching my reasons for putting them aside (which certainly do not amount to decisive objections). I then provide, in the remainder of the paper, my own answer.

2.2 Explanation You could make a case that, although the actual part of a generalization suffices for prediction, the modal part is required for explanation, either because the modal part is essential to the causal aspect of the generalization, which is in turn essential to the generalization's explanatory power (Lewis 1973a, 1986), or because the canons of explanation appeal directly to counterfactual conditionals (Jackson and Pettit 1992; Strevens 2008, §7.3). An explainer, then, has good reason to discriminate in favor of generalizations that support counterfactuals.

I have no objection to this argument, but it is orthogonal to my aim in this paper, to investigate counterfactual support's practical consequences. (You might of course hope that the practical utility of counterfactual support will go some way towards explaining why we care about finding explanations . . .)

2.3 Knowledge Grant the actualist their claim that knowledge of the actual part of a generalization is all that is needed for practical purposes. But contend that knowledge of a generalization's actual part and its counterfactual part are entwined in such a way that knowledge of the actual part cannot be achieved by normal means without acquiring also knowledge of the counterfactual part. This might be for one of two reasons: (a) in coming to know the actual part of a generalization you inevitably or at least typically come to know the counterfactual part, or (b) in order to know the actual part of a generalization, you must first come to know the counterfactual part. The latter possibility would provide the more compelling explanation of our counterfactual concerns: we care about generalizations that support counterfactuals because they are the learnable generalizations, and it is by attending to their modal part that we learn them.

How might you make a case that knowledge of a generalization's counterfactual part is necessary for knowledge of its actual part? Here is one version of the story, chosen for simplicity rather than depth. Suppose that all observed Fs have been Gs. On what grounds can you infer that all Fs are Gs? Arguably, the inference requires you to have some reason for thinking that the G-ness of the observed Fs is not a mere coincidence—that those Fs would have been G under a wide range of counterfactual circumstances. Thus in order to make the inductive inference, you must believe (more or less) that the generalization is, if true, counterfactual-supporting (cf. Dretske 1977 and Armstrong 1983).

This "knowledge explanation" of our concern for counterfactual support is unconvincing, however. First, it does not explain why we continue to care about a generalization's ability to offer support once we have learned that it is true. Is the ladder not now surplus to our needs? Second, the supposition that inductive inference from an observed correlation to a universal generalization can proceed only in the presence of reasons for thinking that the correlation is non-accidental is dubious. Certainly, if we have some reason for thinking that the correlation is a coincidence, the inductive move is defeated. But why not hold that the move is permissible provided that we have no such reason, that is, provided that *as far as we know*, the correlation is not a coincidence? It is hard to see how induction could get off the ground otherwise.

There are other ways you might give a knowledge explanation of our interest in counterfactual support. Indeed, the account I offer later in this paper is a kind of knowledge explanation. But for now, let me put aside the knowledge strategy to look at some alternatives.

2.4 Deliberation The answer I hear most often to the question "Why think counterfactually?" is that counterfactual conditionals play a central role in human deliberation. I think it is plausible that they do play such a role. Even more plausibly—indeed, undeniably—deliberation invokes regularities to predict the consequences of actions. If these regularities are to figure in counterfactual-based reasoning, their consequences must carry over to counterfactual circumstances, which is to say, they must provide counterfactual support. Perhaps we care about a regularity's ability to provide counterfactual support, then, because only supportive regularities are usable in our decision-making.

Let me try to persuade you otherwise; more exactly, let me try to persuade you that this is at best a very weak explanation of our caring about support.

Preamble: counterfactual reasoning should be distinguished from hypothetical reasoning. To represent a state of affairs hypothetically is to represent it without believing it or desiring it, without asserting that it obtains and without being motivated to bring it about. Hypothetical representation and reasoning is essential for deliberation, as a deliberator must represent their possible courses of action and predict the consequences of those actions. For that matter, hypothetical reasoning is essential for any kind of sophisticated learning: to learn inductively, for example, you must represent various competing hypotheses and work out their implications.

The thesis currently under consideration is stronger: it holds that we predict the consequences of our available actions using counterfactual conditionals and not in some other way, such as simple entailment. To evaluate the results of a given action c, that is, we determine for what consequences e the counterfactual conditional $c \mathrel{\Box\!\!\rightarrow} e$ is true, and choose to perform the action for which these consequences are most desirable.

To make this more concrete, let me assume a simple Stalnaker semantics for counterfactual conditionals, on which $c \mathrel{\Box\!\!\rightarrow} e$ is true just in case e is true in the closest possible world where c is true (Stalnaker 1968). (It is assumed, then, that for any way things actually are and any counterfactual antecedent c, there is a unique closest possible world where c occurs; in section 3 I will relax this assumption as recommended by Lewis.) Then to deliberate counterfactually about a possible action is to ask yourself what happens in the closest world where that action is performed, ultimately choosing the action c for which the closest c-world contains the most favorable outcome.

There is a complication: because we are deliberating under uncertainty, we typically do not know which world is actual and so, for any action c, which c-world is closest. We therefore aggregate the possibilities, probabilistically weighted. If you are deciding, for example, whether to perform c or to refrain from performing c (to "perform $\neg c$"), you go through the different ways things might be—the different epistemically possible worlds—and for each find the nearest c-world, evaluating its desirability and adding it to your running total weighted by your subjective probability for the epistemic possibility in question. If the total for performing c is greater than the total for performing $\neg c$—if the probabilistically weighted c-worlds are in aggregate more desirable than the probabilistically weighted $\neg c$-worlds—then and only then should you perform c.

What is wrong with this picture of deliberation? As a description of the way that humans actually reason it is, as I remarked above, plausible if not indisputable. Does that not settle the issue? We humans deliberate using counterfactuals, thus the regularities we invoke during deliberation must be counterfactual-supporting—and that is why we look favorably on regularities that provide support.

This is a possible explanation of our caring about counterfactual support, but it is, I asserted above, a very weak one. To see why, think back to the original motivation for introducing the apparatus of possible world closeness to provide a semantics for counterfactual conditionals, with its convoluted similarity or "closeness" relation (to be described in more detail in section 3.1).

Consider the following simple analysis of a counterfactual of the form *If c had happened, then e would have obtained*: the conditional is true just in case the laws and the background conditions that hold at the time c is counterfactually supposed to occur, together with the proposition that c did in fact occur, entail e. As Goodman (1983) famously pointed out, the analysis cannot be correct, since the laws and background conditions will typically be inconsistent with the supposition that c occurred. When we evaluate the conditional "If this match had been struck five minutes ago, it would have lit", for example, we see that it is clearly and non-trivially true. That cannot be because, when we add the supposition that the match is struck to the laws and background conditions that held five minutes ago, the mix entails that the match lights rather than not lighting, because much of what we know about the laws and background conditions entails that the match was not lit; for example, we know that we had no intention of striking the match at that time.[4] It seems, Goodman wrote, that when we evaluate the counterfactual, we add the supposition of the match's being struck to some subset of the laws and conditions—but what subset? This is Goodman's "cotenability problem".

The possible worlds closeness relation solves the problem (though not quite in Goodman's original terms): find the closest possible world where the match is struck, and see whether the consequent is entailed by the laws and background conditions that hold in *that* world. The laws and conditions will be similar but not identical to the actual-world laws and conditions: the complexities and sophistications of the closeness relation dictate in precisely what respects they differ.

Back to deliberation. My major premise is this: the cotenability problem does not, in practical terms, arise for the deliberator. Why? The deliberator needs to predict the consequences of their available actions. If these actions are genuinely live options, then they are consistent with what the deliberator knows about the laws and background conditions. Thus the deliberator does not face the problem of bracketing some of their background knowledge when making their prediction about a possible action's consequences. They can use it all. If they are a determinist, they know that *something* about the actual background conditions and laws is inconsistent with all but one of their supposedly "live" options, but this knowledge plays no practical role in their deliberation. When adding the proposition that an action is performed to what is known they will face no problem of cotenability—they will have no need to remove something from their representation of the way things are.

What if there are, in the deliberator's mind, a number of possible ways the

4. Goodman does not explicitly consider the analysis in which the background conditions are restricted to those that held at the time of the counterfactual antecedent's putative occurrence; he focuses rather on an analysis in which all background conditions, including present conditions (e.g., the match is currently fresh and unburned) are used. This is, I suppose, because he can see that "rolling back time" will not in general solve the problem. But is suits me rather well to put things in the terms I have, in part because, when deliberating, the putative time of the action is question is *now*, so the time of the antecedent's occurrence and the present time coincide.

world might be, thus (as specified above) a number of different sets of laws and background conditions relative to which they must ask: if I performed *c*, what would happen? The same remark applies. If *c* is a live option in any of these scenarios, then adding it will create no inconsistencies. If it is not a live option, the scenario is irrelevant to deliberation.

If this is correct, then none of the weird and wonderful structure of the closeness relation makes any practical contribution to counterfactual-based deliberation. This is not to say that you cannot deliberate counterfactually; it is to say that nothing that is distinctive or special about counterfactual conditionals, and nothing that is distinctive or special about the cotenability-resolving power of the closeness relation in particular, plays any role in the success of such deliberation. Indeed, you might just as well deliberate using the "add and entail" semantics for conditionals that Goodman showed would not work for counterfactuals in general; that is, you might just as well predict the consequences of an action by asking what the various possible sets of background conditions and laws, when conjoined with the proposition that you perform that action, entail. (You will add the desirability of each set of consequences, weighted by your subjective probability that those laws and background conditions are the actual laws and conditions, to derive the aggregate desirability of the action—which is to say that you will use precisely the same deliberation procedure as a counterfactual deliberator, but using "add and entail" rather than counterfactual conditionals.) Such a decision framework does not require its generalizations to support counterfactuals.

Counterfactual conditionals, and thus counterfactual-supporting generalizations, can be used to deliberate, then, but not in virtue of what makes them uniquely apt for counterfactual thinking. Deliberating, I never need to ask *What would have happened if…*; only *What would happen if….* This makes a deliberation-based explanation of our caring about counterfactual support, or for that matter of our caring about counterfactuals in general, rather unsatisfying.

It is not completely worthless: it can explain why the closeness relation is maximally conservative about the past and not conservative at all about the future (or more exactly, why it takes a *que sera, sera* stance toward the future: what happens is whatever the fundamental laws say happens, given the state of the world at the time of the counterfactual antecedent's occurrence). Neither of these features does anything to resolve the cotenability problem, however; indeed, they simply mirror the structure of the "add and entail" procedure for evaluating conditionals that ignores the cotenability problem altogether.

Suppose that archeologists discover the remains of a new civilization in the life of which a certain distinctive form of sculpture apparently occupied a vital place. At every important cultural center, these sculptures are present in abundance; at every minor site, smaller versions are found. Anthropologists debate the significance of the strange curves of the heads, the eerily flattened features of the faces, the elaborate yet stylized depiction of drapery and jewelry. Then further evidence shows that the statues were widely used to hammer nails. Does that explain why the sculpture had such importance? To be sure, the statues get the job done, but so would a wide range of far simpler implements. It is possible, I suppose, that there is no deeper explanation for the importance of the sculptures than the following: their creators needed something that would hammer nails, and this is what they came up with. The explanation is a feeble one nevertheless. The curves, the flattening, the stylized detail—surely there is some more specific explanation for why these are the way they are?

Likewise, surely there is some more specific explanation for why the possible-worlds closeness relation has the structure that it does, for why we care about counterfactual support relative to this rather than some other conception of closeness?

* * *

A complication: an appeal to counterfactual-based deliberation has often been made by formulators of causal decision theory (Gibbard and Harper

1978; Joyce 1999). Is it possible that counterfactual conditionals, and more particularly the closeness relation, supply a special ingredient, otherwise hard to come by, that causal decision theory needs in order to function properly?

No; an "add and entail" based decision theory can easily be formulated to conform to the requirements of causal decision theory. Two steps must be taken. First, as I have been supposing, you must individuate epistemically possible scenarios—ways the world could be—by their causal structure. Each scenario, that is, must share a set of laws and background conditions, so that the scenarios form what Joyce (1999) calls a K-partition. (To find a correct partition, you must therefore have causal knowledge, a matter that I will discuss in section 2.5.)

Second, in weighting the consequences of performing an action c in a particular epistemically possible causal scenario, you should use your unconditional subjective probability for the scenario, not your subjective probability conditional on c's being performed.

That gives you a decision theory that is, for practical purposes at least, indistinguishable in its recommendations from a counterfactual-based decision theory (Joyce 1999).[5] The ease with which the counterfactual conditionals can be switched out for the "add and entail" conditionals should underline my claim, above, that the closeness-based apparatus for dealing with cotenability is doing no net work in deliberation.[6]

5. I will not make a claim of exact equivalence; that will depend on fine details in the implementation of both frameworks.

6. I say doing no *net* work because counterfactual-based decision theory does in its implementation make use of the closeness relation: in deciding on the value of action c, you are summing not over the epistemically possible causal scenarios that are consistent with any live option, but over possible worlds—complete specifications of matters of fact—that are in many cases, because they specify that c was not performed, inconsistent with c. You need to know which counterfactuals supposing the performance of c are true in such a world, so the details of the closeness relation—the structure that determines its resolution of the cotenability problem—make a difference to your calculations. But it is a difference that ultimately makes no difference: for the reasons given in the main text, the counterfactuals deliver the same answer whatever the details of the closeness relation, an answer determined by the same background conditions and laws invoked by the "add and entail" procedure.

2.5 Causality Causal knowledge matters. It matters, first, because to determine what manipulations of the world will bring about what consequences, we must have some grasp of the world's causal structure (Cartwright 1983). And second, because as explained in the previous section, knowledge of causal relevance is needed to partition the epistemic possibilities into sets of causally equivalent laws and background conditions that are used to deliberate in accord with the principles of causal decision theory. In short, causal knowledge helps us to determine both how to change the world in such and such a way, and whether we ought to change it in that way.

A longstanding and powerful philosophical tradition defines causal notions in terms of counterfactual conditionals (Lewis 1973a; Paul and Hall 2013). Might the practical significance of counterfactual thought, then, be found in the practical significance of causal knowledge?

The story would have to supplemented in some way to explain why we should care whether generalizations offer counterfactual support. (It will be easy to explain why we care whether generalizations are causal; one possibility is therefore to argue that our concern for counterfactual support is a slightly misdirected concern for causality—although as you will see in section 4, a great many counterfactual-supporting generalizations are not purely causal.) Let me, however, put this issue aside.

If understanding causality requires us to think counterfactually, then we

---

To put it another way, the "imaging" operation used to determine the probabilities of counterfactual conditionals and so to implement counterfactual decision theory shuffles epistemic probability around within sets of possible worlds that share the same relevant background conditions and laws—relevant to determining the effects of the action in question, that is—but it never moves probability across the boundaries of such sets (or moves only negligible amounts; see note 5). Since the laws and background conditions fully determine the consequences of an action in a world, this movement of probability has no net effect on the value of the action: for each set the consequences are the same, and so you take all the probabilities that your imaging operation has so painstakingly moved around within the set's boundaries and simply aggregate them, erasing the impact of your rearrangement, thus of that aspect of the closeness relation that resolves cotenability, and ending up with the same results as if you had followed the "add and entail" procedure, which does not pretend to resolve cotenability at all.

have a good explanation why counterfactual reasoning matters. But it is not clear that we do need to be counterfactualists to be causalists. There exist very fine accounts of the nature of causation that make no essential reference to counterfactual conditionals. As an example, let me selflessly cite the theory advocated by Strevens (2008, forthcoming). (Strevens' theory must be supplemented by an account of the relation he calls causal influence; it is possible to give a counterfactual theory of causal influence, but there are good non-counterfactual options as well, such as the theory that Strevens lifts from Dowe (2000).)

One of these non-counterfactual theories might be true. But even if not, the fact that they replicate the counterfactual theories' judgments about matters of causal connection and causal relevance suggests that the counterfactual route is only one of several practically feasible ways of building a causal knower. (I am assuming that such a knower need only to have powerful heuristics for inferring causal facts; they need not grasp the ultimate basis for such facts. Arguably, until we settle our metaphysical differences, we are all heuristic knowers of this sort.)[7]

Again, then, you have a possible explanation for counterfactual thinking, but a rather weak one. What would be far more satisfying—and what I aim to provide—is an explanation that points to the direct practical utility of the finer details of the closeness relation, most importantly the details that play the decisive role in resolving the problem of cotenability.

2.6 Concluding Remarks Let me repeat that the arguments in this section against explaining counterfactual thinking, and the role of counterfactual-supporting generalizations in such thinking in particular, in terms of knowl-

7. Kment (2010) reverses the argument: even if a counterfactual approach to defining causation is incorrect, counterfactual reasoning provides a heuristic so useful for diagnosing causal connections that its value alone explains our tendency to think counterfactually. I reply: there are other equally effective heuristics that do not rely on an elaborate counterfactual logic; why not those?

edge, deliberation, and causation, are far from decisive. In some cases, they are barely sketches of arguments. My aim in advancing them is to convince you that the issues are not yet settled, and so that we should look for new explanations of the importance of counterfactuals and counterfactual support, as I intend to do.

I have not considered every explanation of the importance of counterfactual thinking in the literature. For example, Edgington (2004) suggests (following Adams 1975) that counterfactual conditionals are important in part because they may be used in "counterfactual modus tollens", by which she means this sort of reasoning: "If they were home, the lights would be on; the lights are not on, so they are not home". Or inference to the best explanation: "If they were home, the lights would be on; the lights are on, so (perhaps) they are home".[8] To this proposal, I say the same thing that I said in the discussion of deliberation: such reasoning is useful only when the antecedents are live possibilities, but in that case, the problem of cotenability does not arise, and function of the counterfactual conditionals can be performed by "add and entail" conditionals that do not require of their generalizations anything beyond truth in the actual world. So although the utility of such deliberation may explain some aspects of the rules determining the truth of counterfactual conditionals, they do not explain what is most distinctive about them: the elements of "closeness" that deal with cotenability.

That is a sketch of an answer to Edgington, but there is a limit to what, in this paper, I can usefully do. Let me terminate the negative phase of the argument here, and go on to develop my own explanation of the importance, practically, to us, of counterfactual support.

8. Edgington's observation makes sense of an aspect of the evaluation of counterfactuals—holding to their actual values matters of fact that are determined after, but that are not affected by, the occurrence of the antecedent—that I entirely ignore in this paper. My justification: it is of little importance in the matter of determining a generalization's degree of counterfactual support. Still, this is one of many clues that what I have to say in this paper cannot be, as I will emphasize in my conclusion, the full story about counterfactual thought.

3. The Basis of Counterfactual Support

The actualist's master argument against the importance of counterfactual support is that, because only actual facts matter for practical purposes, a generalization's counterfactual side is without practical significance. But the presupposition of this argument, that a generalization can be divided into actual and modal components, with the facts about counterfactual support confined entirely to the modal part and therefore making no contribution to the actual part, is false: the truthmakers for counterfactual conditionals are themselves in part actual facts with actual-world consequences. To care about a generalization's counterfactual support is in part to care about things that make a difference in actualist terms.

To establish this claim, let me now make good on my earlier promise to explain how generalizations provide counterfactual support. The explanation comes in two parts: an exposition of the received wisdom as to how counterfactuals get their truth values, and a short discussion of the role of generalizations in making counterfactuals true and false, that is, in "supporting" them.

3.1 Truth Conditions for Counterfactual Conditionals What makes a counterfactual conditional true? I will present the version of Lewis's (1973b) answer to this question formulated by Bennett (2003).[9] The aspect of Bennett's truth conditions described here does not cover all counterfactuals; only those that

9. What if you think that Bennett's truthmakers are anything but? For example, what if you think, like Lange (2009), that counterfactuals are made true by primitive subjunctive facts? Or like Fine (2012) that counterfactuals should be defined using an ontology of states rather than possible worlds? Or for that matter, like Lewis (1979) that the closeness relation is not intrinsically time-asymmetric? You ought nevertheless to concede that there are strong correlations between Bennett's putative truthmakers and the real counterfactual truthmakers; thus, you ought to allow that counterfactual conditionals give you information about the Bennett truthmakers. But then, if knowledge of the Bennett truthmakers turns out to be practically useful—as I will argue it is—you should conclude that counterfactual conditionals give you practically useful information. In short: you might object to the narrative framework, but you may still accept the moral of my story.

might be called ordinary or "standard" (Lewis's term) counterfactuals. It is in virtue of precisely these ordinary counterfactuals, however, that scientific generalizations offer what counterfactual support they do; the restriction will not, then, have a significant impact on the argument.[10]

On Bennett's version of the story, a counterfactual conditional of the form *If c had occurred at t, then e would have occurred* is true if *e* holds in all (or nearly all) of what I will call the "evaluation worlds" for *c*.[11] The evaluation worlds are determined as follows. First, gather together all the possible worlds (or possible scenarios, models, etc.) satisfying the following description:

1. The world is identical to the actual world until shortly before *t*.

2. At that point events deviate conservatively from the actual course of events, so as to bring about *c* at time *t*.

3. Thereafter, events are determined by the actual world's laws of nature acting on the state of the world at time *t*.

The deviation in question will require a "small miracle"—a violation of the actual laws of nature—if the laws are deterministic; if they are indeterministic, it may require only that an indeterministic process shortly before *t* yields a different outcome than occurred in the actual world.

The question of what makes for a conservative deviation is complex, but for Bennett, adhering more or less to Lewis's story, a deviation is conservative to the extent that it minimizes: (a) violations of the actual laws of nature, (b) in

10. Three kinds of counterfactuals not covered by Bennett's account are: (a) "backtracking" counterfactuals, (b) counterfactuals whose antecedents envisage large-scale changes in the world, such as "If Caesar had been in charge in Korea . . . " or "If I were a dog . . . ", (c) counterfactuals in which conversational context plays an important role in determining the relevant "closeness" relation, many of which are of types (a) and (b).

11. Normally the nature of *c* implies the relevant time *t*, since singular events are individuated in part by their time of occurrence; here I distinguish *t* and *c* for expository convenience. More generally, in what follows I assume that antecedents are events counterfactually supposed to occur at a particular time; when this assumption does not hold, the "closeness" metric will have to do the additional work of determining appropriate events and times.

an indeterministic world, the occurrence of highly improbable events, (c) the duration of the period between its departure from actuality and *t*, and (d) in the period between its departure from actuality and *t*, the degree of difference between the deviation and actuality with respect to particular matters of fact. A deviation is conservative, then, if it departs discretely from reality with few discernible side effects, and without bending the fundamental laws further than is absolutely necessary.

The second step in determining the evaluation worlds is to select from the above set those that deviate from actuality most conservatively. These are what Lewis and others call the "closest possible worlds" in which *c* occurs at *t* (or closest possible world, singular, in Stalnaker's case).

3.2 The Basis of Counterfactual Support Apply Bennett's truth conditions to one of the ordinary counterfactuals for which the raven color generalization—*All ravens are black*—is said to provide support:

> If these two ravens were to have mated (at time *t*), their offspring would have been black.

To evaluate the conditional, look to worlds where the most conservative deviations from actuality before *t* result in a mating—such as worlds where the female raven becomes ready to breed a day earlier than in actuality or where a storm blows the male raven far from its home territory.

What are these worlds like? I assume that ravens are black for the following schematic reason: normal ravens have a complex of physical properties *P* that, in normal conditions, both colors their feathers black and gets itself passed on to future generations of ravens. (*P* is not a single physiological mechanism, of course, but a set of mechanisms.)

Under these assumptions, an evaluation world for the raven mating will have the following character.

1. Up until shortly before *t*, it will be identical to the actual world. Thus

as in the actual world, the parents have the property *P* and normal conditions hold (so I assume—or else the conditional is false).

2. Shortly before *t*, events deviate from actuality so as to bring about the mating. Because the deviation is conservative, it has few side effects; in particular, it does not interfere with the *P*-hood of the parents or the fact that normal conditions obtain.

3. After *t*, the actual laws of nature operate on the state of the world at *t*; since normal conditions hold and the raven parents have *P*, their offspring also has *P* and so is black.

All of this follows immediately from the definition of the evaluation worlds except for the claim in (2) that the deviation will not undermine the parents' *P*-hood or normal conditions. Let me explain why the claim is true. I will focus on *P*-hood; the same treatment will apply to the normal conditions.

The most conservative deviations that bring about mating will not undermine the parents' *P*-hood for two reasons. First, *P*-hood has what I will call *physical inertia*: it tends to persist unless something actively works to remove it—a result of the raven body's many mechanisms for self-maintenance. Second, the antecedent of the conditional, that is, the mating of the ravens, can be brought about in ways that avoid undermining *P* without making conservatism-depleting sacrifices elsewhere. Since the most conservative deviations by definition avoid making changes to actuality that they do not have to make, they will avoid undermining *P* in particular; thus, the raven parents will retain their *P*-hood in the evaluation worlds.

Why is it possible to bring about the mating without undermining *P*? Because the mating and the ravens' *P*-hood are *physically and causally separable*. Physical separability means that there is no overlap in the realization of mating and *P* in worlds that are biologically like our own, so that a change to a raven's mating history does not mandate a change to the physical facts that constitute *P*-hood. Causal separability means that there is no overlap in the realization of mating behavior and the creating or sustaining causes of P-hood. Where there is separability between two properties, there is typically the prospect of manipulating one without thereby changing the other. (More could be said here, but my aim is to elucidate the major factors at work in explaining why P-hood is left intact, not the more complex task of giving necessary and sufficient conditions for non-undermining.)

To sum up, the facts that play a role in making the mating conditional true are the following (omitting for clarity's sake mention of "normal conditions"):

1. The parents in question have P,
2. P-hood has physical inertia, that is, a tendency to persist,
3. P-hood is physically and causally separable from the mating, and
4. By way of the fundamental laws of physics, P-hood causes itself to be transmitted to the next generation, where it then causes blackness.

Observe that each of these is either an actual fact—the sort of fact that actualists are happy to allow makes a practical difference to our lives—or, like the fundamental laws, has a substantial actual part, hence many implications for the actual facts.

3.3 The Nature of Counterfactual Support In what sense does the raven color generalization "support" the mating counterfactual? As follows: the truthmakers for the mating conditional are simply the truthmakers for the generalization itself, or are "particularized" versions of those truthmakers.

What are the truthmakers for the generalization, then? This question merits some serious consideration, but let me keep things brief by putting a proposal on the table without any preliminary discussion. The raven color generalization's truthmakers are, I suggest, the following:[12]

12. Here and in the remainder of this section, I ignore the question whether the generalization is implicitly qualified by "normality" riders or any such thing; the issue will return in a certain guise in section 4.

1. The actual P-hood of most or all actual ravens,
2. The physical inertia of P-hood, that is, its tendency to persist,
3. The causal and physical separability of P-hood from a wide range of relevant counterfactual antecedents, and
4. The aspects of the fundamental laws of physics in virtue of which P-hood causes blackness and transmits itself to the next generation.

Together these entail what you might think of as both the actual and the modal components of the raven color generalization:

> *Actual component:* All actual ravens are actually black.
>
> *Modal component:* Actual ravens in many counterfactual circumstances are black, as are many counterfactual ravens (such as ravens produced by counterfactual matings).

This is all the justification I will give for my list of truthmakers, since the question whether the list is entirely correct, and in particular whether it is complete, is secondary to this paper's principal concerns.[13]

It is clear that the truthmakers for the raven color generalization include the truthmakers for the mating counterfactual (compare the two lists); further, each of the truthmakers for the generalization plays a role in making the counterfactual true. (In the case of (3), it is a "particularized" version of the generalization's truthmaker that makes the counterfactual true—namely, the separability of P-hood from the particular counterfactual antecedent in question rather than from a range of such antecedents.) My proposal, then, is that the raven color generalization "supports counterfactuals" means that the truthmakers for the generalization coincide with the truthmakers for each of

13. For further discussion of this view of what makes a causal generalization true, see Strevens (2008), §7.6.

a wide range of ordinary counterfactual conditionals just as described. Some remarks on this definition of counterfactual support.

First, a generalization like *Ravens are black* has exceptions. Some ravens—leucistic specimens, for example—have a defect in their coloration mechanism and so lack some crucial component of the all-important property *P*, as a result of which they are light gray or white. For much the same reason, the raven generalization will not support every counterfactual. The genetic makeup of two normal ravens might be such that their offspring has considerable chance of being leucistic; in that case, it is not true that if the ravens had mated, their offspring would be black. Thus my rather loose definition: for such a generalization to provide counterfactual support it must support a "wide range" of counterfactuals. In the special sciences, providing counterfactual support is a matter of degree.[14]

Second, there is a distinction to be drawn between a generalization's support of counterfactuals and a generalization's robustness under counterfactual suppositions. To see that they come apart: define a raven* as a raven with a yearpoint in *Y* (where *Y* is, as before, the set of all actual ravens' yearpoints). *Ravens* have Y-yearpoints* is a necessary truth, and so absolutely robust under counterfactual supposition. But it no more supports counterfactuals than its ultra-fragile counterpart, *Ravens have Y-yearpoints*: if this particular raven had made a left turn at Albuquerque, it would not have had a *Y*-yearpoint (and so it would not have been a raven*). Nor is it of any greater interest than its fragile counterpart; it is counterfactual support that matters to us, then, not robustness under counterfactual supposition.

Third, that for a generalization of the form *All Fs are G*, a wide range of corresponding counterfactuals are true, is necessary but not sufficient for

14. When we say that a fundamental law supports counterfactuals, we mean that it supports any counterfactual with a physically possible antecedent, Roberts (2008) and Lange (2009) persuasively argue. Special science generalizations by contrast will typically fail to support some counterfactuals with antecedents that are perfectly possible by the special science's own lights.

the generalization to be counterfactual-supporting, because the truthmakers for the generalization may not bear the right relation to the truthmakers for the counterfactual. To see this: Let w be the weight of the fattest raven ever, plus a little bit. It is then true that *All ravens weigh less than* w. Further, a range of corresponding counterfactuals are true, such as *If this (skinny) raven had eaten a bacon double cheeseburger, it would still have weighed less than* w. Yet the generalization does not support the counterfactuals. Why not? The counterfactuals are supported by a certain generalization about the physiology of ravens, which concerns the relation between calorie intake and weight gain—call it the metabolic generalization. The "*w*" generalization, however, contains strictly more content than the metabolic generalization, namely, whatever must be added to the metabolic generalization to entail that no raven weighs more than w.[15] This additional content—this additional truthmaker—plays no part in securing the truth of the counterfactuals; thus, the "*w*" generalization does not, on definition that requires all truthmakers to play a part, support the counterfactuals.

The actualist's dismissal of the importance of the counterfactual realm has as its major premise the proposition that a generalization's modal component has no implications for actuality that are not already present in the actual component. But this, I hope you can now see, is incorrect. To care about the modal side of a counterfactual-supporting generalization such as *All ravens are black* is to care about items (1) through (4) above: the P-hood of ravens, the physical inertia and separability of P-hood, and the causal or nomological consequences of P-hood. These are facts with actual consequences, and they are not entailed by the actual blackness of actual ravens.

What, then, is the practical significance of these facts? How are the actual-world consequences of (say) *Ravens are black* expanded by appreciating not only that it entails the blackness of actual ravens, but also by appreciating

15. I suspect that this additional something is simply the fact that no raven weighs more than w, but its identity is unimportant here.

facts about the separability and physical inertia of P-hood?

4. Practical Counterfactualism

Knowing that all actual ravens are actually black gives you enormous predictive power in the matter of raven color. How could such power possibly be enhanced?

But not all ravens are black. Some have been painted white. Some are leucistic; some are albinos. Thus you cannot know that all actual ravens are actually black—it is simply not true. What you come to know instead, when you first learn something about the color of ravens, is that most ravens are black. This is useful knowledge, but it would be more useful still if you could add to it a specification of the conditions in which exceptions can be expected. It would be more useful still, that is, to augment your knowledge about actual raven blackness so that it did not take the form "Most ravens are black" but rather something like "In conditions Z, all ravens are black" for some fine-grained specification of conditions Z that pinpoints just those circumstances when a raven is guaranteed to be black. Of course, knowledge of this latter sort is unlikely to be fully achieved (and in an indeterministic world, may be impossible), but there is a significant practical payoff to having it even in part. The actualist will happily endorse this claim, as it asserts the utility, in principle, of fine-grained knowledge of actual correlations.

I will argue for the following thesis: the counterfactual-supporting generalizations are those for which there exists a compact and (under normal circumstances) knowable basis determining the content of Z, that is, determining the fine-grained pattern of correlation between the properties associated by the generalization. We are interested in generalizations that offer counterfactual support, then, because it is these generalizations, and only these generalizations, concerning which we can learn not just a loose statistical correlation but a fine-grained, near-universal association. The difference has not to do with the actuality of what we seek to know, but with the degree of detail with which that actuality is knowable. (As you will soon see, there is more to knowledge of fine-grained variation than knowledge of *Z*, but let the part stand for the whole for the time being.)

The argument will come in two parts. The first part concerns generalizations of the form *In conditions Z, Fs are G* that hold in virtue of a causal mechanism by which *Z*, *F*, and (possibly) some other properties bring about *G*.

The second part concerns generalizations that hold in virtue of both a causal mechanism and a correlation between the high-level property that appears in the antecedent of the generalization and another property that does the actual causing. The canonical case is a generalization of the form *In conditions Z, Fs are G* that holds in virtue of (a) the fact that most or all *F*s have some property or property complex *P*, and (b) a causal mechanism by which *Z*, *P*, and (possibly) some other properties bring about *G*.

An example of the second sort is the raven color generalization, since blackness in ravens is not caused by ravenhood per se but rather by certain physical properties of ravens—namely, the properties that make up the raven coloration mechanism, the presence of which is neither necessary nor sufficient for ravenhood. (The raven coloration mechanism is not necessary for ravenhood because leucistic ravens do not have it and, more generally, because ravens could have evolved to be some other color; it is not sufficient because other birds might have been, and in the case of raven congeners such as carrion crows plausibly are, black in virtue of the same mechanism.)

Consider the generalization *Fire burns*. I will suppose that it is true wholly in virtue of a certain causal mechanism in which fire plays the principal (though not the only) role; it belongs to the first class of cases, then.

Knowledge of the fire generalization is rather useful. I can use it to avoid injuring myself or to injure my enemies. I can also use it to warm or cook food (if I know the straightforward connection between burning, warming, and cooking), and to warm myself. To truly exploit the power of fire, however, the simple knowledge that fire burns hardly suffices. If I want to use fire to cook food or stay warm without injuring myself, or if I need to fish some dropped object out of the fire safely, I require a rather more fine-grained knowledge of the connection between fire and heat. I would do well, for example, to understand that at comparable distances it is much hotter above, than to one side of, the flames. It will be useful to know that exposure that in the long term would cause serious burns can be tolerated without major damage in the short term (when manipulating the fire, say). I would, in other words, benefit greatly from coming to know a complex generalization relating fire and burning of which *Fire burns* is a rather crude summary.

How will I come to know and to represent this complex generalization? In principle, I could learn everything I need to know through brute statistical testing, and store that knowledge in a table of correlations. It would be far more efficient, however, to come to some understanding of the mechanism by which fire burns, and to store the knowledge in a causal theory—a representation of certain aspects of the mechanics of burning. This understanding might be relatively shallow and incomplete. It might consist entirely in some of the following precepts, for example: It is the accumulation of heat that burns (so burning increases with exposure time). The rate of heat accumulation falls off with distance. Heat radiates in all directions, but most of all, it rises.[16] This "theory of heat" could be enriched by further, deeper theory: you might realize, for example, that heat rises because hot air rises. But such information is supplementary; the principles I have described are quite useful without it. More causal understanding is always good (at least in the quantities that were available in pre-scientific times), but even a little goes a long way.[17]

Strevens (2007) elaborates the case for the practical value of small amounts of information about underlying mechanisms. Counterfactual thinking, I should emphasize, is in no way needed to realize this value, which inheres

16. The pre-scientific notion of heat may not fully distinguish heat and temperature (Wiser and Carey 1983).

17. This is just as well; Rozenblit and Keil (2002) show how shallow most people's understanding of many mechanisms can be.

in the implications that facts about mechanisms have for actual variation. (Indeed, one of the purposes of Strevens (2007) is to explain why we think causally without appealing to causal, counterfactual, or other modal facts.) I will not add anything further to the argument here, except to press the importance of one class of cases. I speculate that a considerable number of the generalizations worth knowing are social and psychological generalizations, for example, *If you have more material possessions than other people, they will tend to envy you*. The riders on these generalizations—the *Z* that makes *In conditions Z, if you have more material possessions than other people, they will envy you* exceptionlessly rather than "statistically" true—will be tremendously complex; understanding psychological and social mechanisms will give you a good chance, and perhaps your only chance, of grasping a useful proportion of this complexity.

Let me now turn to generalizations that are true in part because of a correlation between a high-level property specified in the antecedent and an underlying causally efficacious property, such as *Ravens are black* or *Roasted raven is good to eat*. Such generalizations are useful to us because the efficacious property in question is typically hard to detect while the high-level property is readily observable (Strevens manuscript). From the presence of the high-level property, then, we infer the presence of the relevant causal factors. Our ability to make this inference will of course be greatly and fruitfully enhanced if we understand precisely the circumstances under which causal properties are likely to come along with the high-level properties.

To put things schematically, suppose that the generalization in question is *Fs are G*, and it is true in virtue of a mechanism largely composed of certain unobservable properties *P* of the *F*s. Our ability to use the generalization will improve the better we understand the circumstances under which *F*s have *P* and the circumstances under which *P* causes *G*. The latter question was discussed above; in what follows, the aim is to say something more about our understanding of the connection between *F*-ness and *P*-hood.

One source of understanding is knowledge of the mechanism that produces *P* (if there is such a mechanism). For example, in the case of raven blackness, knowledge of the causal process by which ravens come to have the blackness-producing property *P*—a process involving raven reproduction and development—will tell you quite a bit about when a raven is likely to have *P*.

Such knowledge may, however, be difficult to come by, especially in the case where *P* itself is unobservable.[18] Consider such a case—a case where you know nothing about the origins of *P*-hood. Because the great majority of *F*s have *P*, you may infer, of any particular *F*, that the *P*-producing mechanism has done its work, and so that the *F* in question has *P*. Of course, this assumption takes advantage of no knowledge of fine-grained variation, but since you have none, it is the best you can do. Crucially, the story does not end there. Though the *F* at hand has (so you assume) *P*, there is in many cases still the potential for its *P*-hood to be disrupted—to disappear—before it can bring about *G*. This is especially true if you or someone else is using the *F* for some purpose, since the *F* will then quite likely be taken out of its usual environment.

Consider *Roasted ravens are good to eat*, for example. Presumably ravens have some intrinsic property *P* in virtue of which this generalization is true (something to do with the quality of the meat, the lack of toxins, and so on). If you wish to take advantage of the correlation, you must find a raven, then kill, dress, and cook it. These processes may, for all you know, interfere with one of the properties of the raven that make it good to eat, that is, they may undermine the raven's *P*-hood. (We are supposing *faute de mieux* that the raven you found had *P* to begin with.) You would, then, like to know what kinds of things you can and cannot do to the raven without interfering with *P*.

18. It is certainly not impossible to come by. By recognizing signs of, say, a developmental disorder, you may infer that some particular specimen of ravenhood is less likely than most to have the full complement of normal raven properties.

The facts in question will typically be highly complex. A bad way to learn and represent them would be to compile a list of everything that might disrupt *P*—bad because the list would be heterogenous and long, perhaps infinite (though it might of course be worth learning by rote some especially important disrupters). A much better strategy is possible if it is assumed that *P* has a certain degree of physical inertia. That strategy is as follows. Observe or infer what kinds of things *P* is separable from, physically and causally. Then use your physical knowledge, most importantly your causal knowledge, to determine for any particular scenario whether *P* is likely to be disrupted. Knowing the (rather obvious) fact that a raven's palatability inheres in its muscle fiber and not its brains, for example, you might be more careful, in your killing and dressing, with former than the latter.

Let me put this proposal in general terms. Given a regularity *Fs are G*, to learn and represent the fine-grained patterns of variation between properties *F* and *G*, where *F* is not itself a cause of *G*-ness, proceed as follows.

1. Assume that most or all *F*s possess an *G*-causing property *P*, perhaps unobservable.

2. Assume also that *P* possesses physical inertia. (Where this and the previous assumption cannot be made, the method is of no use.)

3. Learn as much as you can about what *P* is physically and causally separable from. This will tell you which manipulations of a given *F* are more likely, and which are less likely, to undermine its *P*-hood (regardless of whether the manipulations are performed by you, another agent, or nature).

4. Learn as much as you can about the mechanism by which *P* causes *G*. This will give you a better appreciation of, first, the conditions under which *P* is likely to cause *G*, and second, the quantitative, spatiotemporal and other relations between *P* and *G* that are likely to result from such a causal process.

As noted above, an alternative to, or better a supplement to, steps (1) through (3) is to learn as much as you can about the causal mechanism and other facts responsible for the *P*-hood of *F*s. This information of course partly overlaps the information in step (3).

To implement this procedure, you must attend to precisely the properties in virtue of which the generalization in question—*Fs are G*—supports ordinary counterfactuals. Further, the procedure works only if the generalization *does* offer counterfactual support (and it works better, providing more information about fine-grained variation, the more support is offered). This immediately suggests an explanation of why we care about counterfactual support. We care *that* generalizations offer counterfactual support because it is only when they do that we can (in practical terms) learn the fine-grained variation in the pattern in question. And we care *how* generalizations offer counterfactual support because the same elements that explain counterfactual support instruct us as to the nature of that fine-grained variation.

Let me take a closer look at the features of and the lacunae in this explanation.

5. What Is Explained?

The generalizations that support ordinary counterfactuals are precisely those for which a compact, knowable basis for actual fine-grained variation exists. Further, contemplation of the ways that such a generalization offers counterfactual support reveals the patterns of fine-grained variation, because the basis of actual fine-grained variation is contained in the basis for support.

This explains, I propose, our interest in generalizations' ability to offer counterfactual support: our concern with a generalization's implications for practically irrelevant alternative histories directs our attention to the fine-grained variation in actual matters of fact on which our aspirations and achievements depend.

The story assumes, but does not account for, a habit of counterfactual thought—a practice of thinking about unrealized possibilities using counterfactual conditionals. Given the existence of this cognitive quirk, and given the nature of the closeness metric that determines the truthmakers for counterfactual conditionals, an additional habit—attending to counterfactual support—will have supreme practical value.

Could this practical value explain counterfactual thought itself? It seems unlikely: there are some rather more straightforward ways that our attention might be directed toward the basis of fine-grained variation. We might, for example, contemplate not counterfactual but future subjunctive conditionals that hinge on the same basis, or we might simply find ourselves fascinated by these properties in their own right. (We are already quite interested in causal laws, of course.) Besides, what I have to say applies only to ordinary counterfactuals.

My explanation makes more sense, then, if you suppose that a pre-existing tendency to think counterfactually was exploited, by inculcating a tendency to care about counterfactual support, to make us experts in fine-grained variation, rather than as being originally constructed primarily for that purpose. In that case, I have explained why we value generalizations' counterfactual support but not counterfactual thought itself.

Perhaps this is enough for one paper. But still, it seems that more might be said: is it just an extraordinary coincidence that there is such a close match between the basis of counterfactual support and the basis of actual fine-grained variation?

I think not. I think that there is a real possibility that the pursuit of knowledge of fine-grained variation explains the structure of the counterfactual closeness relation itself. It does not explain counterfactual thought—that goal remains beyond the bounds of this paper—but it explains how such thought became entrained to a particular way of determining which worlds matter for the purposes of evaluating ordinary counterfactual conditionals, that is, a particular way of resolving the problem of cotenability.

Here is the explanation I have in mind. As before, I conjecture, we have for independent reasons a tendency to think counterfactually: to contemplate unrealized possibilities, to ask "What if?" about scenarios that are now foreclosed. But the role of this speculation about matters contrary to fact places constraints on the relation's structure that are quite weak.

(What role might that be? Perhaps imaginative retrospection about how things might have worked out better for me than in fact they did, to stimulate more careful and comprehensive planning next time around. Or perhaps deliberation, remembering that successful planning and choice of action does not require of the closeness relation any particular solution to the cotenability problem, but only conservatism with respect to the past and openness with respect to the future.)

Given this cognitive backdrop, there would be an advantage to augmenting our conception of closeness so as to take into account the facts that determine fine-grained variation—inertia, separability, and so on—because time spent thinking about counterfactual what-ifs, and about the range of what-ifs supported by practically important generalizations in particular, would then be time spent attending to facts essential in anticipating fine-grained variation. We would thus commit ourselves to a particular way of solving the cotenability problem, of deciding how, given a counterfactual antecedent, to build a consistent possible world in which that antecedent occurs.

This narrative is rather loose, of course. What comes first, a concern with support or the precisification of the closeness relation? And by what process—evolution? learning?—is advantage converted into actuality? But these questions are in any case perhaps best left to cognitive anthropologists. I am happy simply to suppose that "there's a divinity that shapes our ends, rough-hew them how we will."[19]

* * *

19. *Hamlet*, Act 5, Scene 2.

With my explanatory claims clarified, I can address at last a concern foreshadowed in section 3.2: the basis of actual fine-grained variation is not identical, but is rather a proper part of, the basis for counterfactual support. More specifically, the counterfactual support basis includes not only the basis of actual fine-grained variation but a "modal halo".

To see the modal halo, compare the basis for the counterfactual support offered by the raven generalization (section 3.3) with the basis for the actual fine-grained variation in raven blackness (section 4). For fine-grained variation, what matters is separability in the actual world, whereas for counterfactual support, separability in close possible worlds matters too. The same goes for inertia and also of course for the fundamental laws of physics: their actual-world component is sufficient to predict fine-grained variation, but they must have some validity in nearby possible worlds if counterfactuals are to be supported.

Is the halo a problem for the explanations proposed above? The first explanation assumes, without purporting to explain, the existence of a tendency to counterfactual thought using conditionals defined in terms of the familiar closeness relation. It then accounts for our caring about counterfactual support by noting that attention to a generalization's ability to support counterfactuals is also attention to the basis of that generalization's fine-grained variation. This requires only that the variation basis be a subset of the support basis; the halo, then, presents no difficulty.

What of the second explanation? It also assumes without purporting to explain a tendency to counterfactual thought. But it attempts to account for the structure, or at least the details of the structure, of the familiar closeness relation, by pointing to the importance of the elements that determine closeness for understanding actual variation. Here you might think that the halo poses a problem; I will argue that, on the contrary, this story accounts for the halo—it explains why, though the modal element of the support basis plays no role in understanding fine-grained variation, we care about it nevertheless precisely because we care about fine-grained variation.

I will proceed in two steps. First, I suggest that modal aspects of separability and inertia are due entirely to their actual world bases along with the modal element of the fundamental laws. The inertia of the raven coloration mechanism, for example, is due to facts about the actual world—about the mechanism and the physiology of ravens more generally—plus the fact that the laws that determine the effects of raven physiology hold in the kinds of closest possible worlds that are relevant to the raven generalization's provision of counterfactual support. The same is true for separability: what carries it over to neighboring possible worlds is the laws' own modal halo. The element of counterfactual support that goes beyond the basis of fine-grained variation, then, is a single magical, modal ingredient: the modal aspect of laws, or in other words, the aspect of laws in virtue of which they hold in nearby possible worlds in addition to the actual world.

Second, I suggest that this ingredient, the modal halo of the fundamental laws, is derived from the laws' role in the definition of closeness, rather than being something that we care about independently and so build into closeness. That might seem wrong: are the laws not nomologically necessary in virtue of their lawhood? Is this not why they hold in close possible worlds?

No: the close possible worlds relevant to counterfactual conditionals are frequently, if not always, nomologically impossible, because of antecedent-enabling "small miracles" that involve the violation of the actual fundamental laws of nature. The laws hold in such worlds, insofar as they do, not because of the natural modal dominion proper to lawhood, but because we have defined closeness in a certain way. What explains the definition of closeness, then, explains the modal halo of lawhood, inertia, and separability that distinguishes the basis of a generalization's counterfactual support from the basis of its actual fine-grained variation. If the practical importance of grasping fine-grained variation explains why we build the basis for such variation into the closeness relation, then it thereby also explains the modal aspect of the support basis—for that modal aspect derives entirely from the variation basis and the structural significance, in counterfactual thinking, of closeness.

Let me rehearse the resulting explanation of the closeness relation and the halo. We have a practical interest in the basis of the world's fine-grained variation. Because of that interest (though no particular process is described to flesh out this "because"), we privilege a closeness relation that is built on that basis, and on nothing else, that is, a closeness relation designed to carry over just that basis to those possible worlds that are used to evaluate counterfactual conditionals. This very carrying over bestows a modal halo on the basis, in the form of the (suitably circumscribed) truth of the laws and of their consequences—such as stability and inertia—in these worlds. The modal halo does not motivate our caring about closeness and thus the modal elements of the basis of counterfactual support, then; it is, rather, a side effect of shining a spotlight on the basis of fine-grained variation using the counterfactual apparatus.

And what is not explained? Plenty. Why do we think about foreclosed possibilities at all? Why do we do so using conditionals with a certain broad logical structure? Why do we exalt the basis of actual fine-grained variation by filling out the structure of the closeness relation in a certain way, rather than by some other cognitive strategy that has nothing to do with counterfactual thought? All this goes unanswered. Perhaps the answers, when they come, will add significantly to the story I have offered about the closeness relation and the importance of counterfactual support. Meanwhile, I hope I have made a useful start.

Acknowledgments

My thanks to: Alison Fernandes, Boris Kment, Marc Lange, John Roberts, and audiences at the Chapel Hill Colloquium in Philosophy and the University of Stockholm.

References

Adams, E. W. (1975). *The Logic of Conditionals*. D. Reidel, Dordrecht.

Armstrong, D. M. (1983). *What Is a Law of Nature?* Cambridge University Press, Cambridge.

Bennett, J. (2003). *A Philosophical Guide to Conditionals*. Oxford University Press, Oxford.

Cartwright, N. (1983). Causal laws and effective strategies. In *How the Laws of Physics Lie*. Oxford University Press, Oxford.

Dowe, P. (2000). *Physical Causation*. Cambridge University Press, Cambridge.

Dretske, F. (1977). Laws of nature. *Philosophy of Science* 44:248–268.

Edgington, D. (2004). Counterfactuals and the benefit of hindsight. In P. Dowe and P. Noordhof (eds.), *Cause and Chance: Causation in an Indeterministic World*. Routledge, London.

Fine, K. (2012). Counterfactuals without possible worlds. *Journal of Philosophy* 109:221–246.

Gibbard, A. and W. Harper. (1978). Counterfactuals and two kinds of expected utility. In C. A. Hooker, J. J. Leach, and E. F. McClennen (eds.), *Foundations and Applications of Decision Theory*, volume 1. D. Reidel, Dordrecht.

Goodman, N. (1983). *Fact, Fiction, and Forecast*. Fourth edition. Harvard University Press, Cambridge, MA.

Jackson, F. and P. Pettit. (1992). In defense of explanatory ecumenism. *Economics and Philosophy* 8:1–21.

Joyce, J. (1999). *The Foundations of Causal Decision Theory*. Cambridge University Press, Cambridge.

Kment, B. (2010). Causation: Determination and difference-making. *Noûs* 44:80–111.

Lange, M. (2009). *Laws and Lawmakers: Science, Metaphysics, and the Laws of Nature*. Oxford University Press, Oxford.

Lewis, D. (1973a). Causation. *Journal of Philosophy* 70:556–567.

———. (1973b). *Counterfactuals*. Harvard University Press, Cambridge, MA.

———. (1979). Counterfactual dependence and time's arrow. *Noûs* 13:455–476.

———. (1986). Causal explanation. In *Philosophical Papers*, volume 2, pp. 214–240. Oxford University Press, Oxford.

Paul, L. A. and N. Hall. (2013). *Causation: A User's Guide*. Oxford University Press, Oxford.

Roberts, J. T. (2008). *The Law-Governed Universe*. Oxford University Press, Oxford.

Rozenblit, L. R. and F. C. Keil. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science* 26:521–562.

Stalnaker, R. (1968). A theory of conditionals. In N. Rescher (ed.), *Studies in Logical Theory*. Blackwell, Oxford.

Strevens, M. (2007). Why represent causal relations? In A. Gopnik and L. Schulz (eds.), *Causal Learning: Psychology, Philosophy, Computation*. Oxford University Press, New York.

———. (2008). *Depth: An Account of Scientific Explanation*. Harvard University Press, Cambridge, MA.

———. (2012). Ceteris paribus hedges: Causal voodoo that works. *Journal of Philosophy* 109:652–675.

———. (Forthcoming). Causality reunified. *Erkenntnis*.

———. (Manuscript). High-level exceptions explained.

Wiser, M. and S. Carey. (1983). When heat and temperature were one. In D. Gentner and A. Stevens (eds.), *Mental Models*. Lawrence Erlbaum, Hillsdale, NJ.