Towards a Unified Normativist Theory of the Natural Modalities

John T. Roberts (johnroberts@unc.edu)

Draft for UT-Knoxville Seminar, Feb. 20, 2014

1. Introduction

In this paper I'm going to make a tentative start at a new way of theorizing about the *natural modalities* – those modally-inflected notions including laws of nature, nomic necessity, causation, counterfactuals, chance, dispositions, and powers, which play such prominent roles in scientific explanation. I think – or anyway, I hope –that the approach I'm going to start developing here can eventually produce a unified theory of all these natural modalities. But here I will focus on causation, counterfactuals, and nomic necessity.

I will pursue a *normativist* theory of these modal notions. Roughly speaking, that means that I will propose that we understand counterfactuals and discourse about laws and causes as not in the business of describing, but rather in the business of prescribing and evaluating. There are a lot of different ways of understand what this really comes to. You might say that statements about the natural modalities don't have truth values, because they don't state facts, but rather give instructions. Alternatively, you might think of such statements as having truth values, and say that their truthmakers are to be found among facts of a very special sort – normative or evaluative facts. Another possibility is to adopt a deflationary account of truth and a quasi-realist account of normative and evaluative discourse, and to say that statements about the natural modalities really are true or false in virtue of how things are – but to then insist that this needn't entail any ontological heavy weather. No doubt there are other options too. For purposes of this paper, I don't much care how you make your choice among these options in philosophy of language. What matters here is the (metaphysical?) idea that how it is with the natural modalities is to be ultimately explained in terms of how it is with what's good and what we ought to do. So, insofar as we ought to do things only insofar as we have some freedom, or at least genuine agency, the account of the natural

1

modalities I am aiming for could not possibly undercut the notions of freedom and agency, no matter what form the laws and the causal relations take. (Unless, that is, my account ends up undermining its own foundations – let's ignore that possibility.)

A normative approach to modality is hardly a new idea. In the last century, a number of philosophers proposed normativist accounts of causation, laws, and/or counterfactuals, including Gilbert Ryle (1950/1971), Wilfird Sellars (1954), Simon Blackburn(1986), and Robert Brandom (2008). More recently, Amie Thomasson (2007) has defended an account of metaphysical necessity according to which claims of metaphysical necessity express semantic norms.

The idea that Ryle and Brandom started from was that claims about laws express *proprieties of inference*. I don't think *this* idea has much hope. The basic problem can be put like this: A law makes it okay for us to draw an inference supported by that law only if we are justified in believing that it *is* a law – but even an accidentally true generalization can justify our drawing an inference if we are justified in believing in it. So for example, suppose it's a law that all copper conducts electricity, and it's just a crummy brute fact that all of the objects on Alex's workbench right now are made of copper. If you're lucky enough to know both of these facts, then from "x is made of copper," you can reasonably infer "x conducts electricity, but you can will equal justice infer from "x is on Alex's workbench right now" to "x is made of copper" (and thence to "x conducts electricity"). There just is no asymmetry here between the law, which is counterfactually robust, and the crummy brute fact, which is very counterfactually fragile.

So it isn't norms of *justified inference* that matter for the natural modalities; if a normativist account of laws and so forth is forthcoming, it will have to be based on some other sort of norm. What other sort of norm will do the trick? The account I present will be based on *methods* of two sorts: Practical methods, i.e. methods for getting things done, and observational methods, i.e. methods for finding things out on the basis of empirical input of some sort. Methods have to do with normatively in two different ways: First, each method comes with a set of standards that determine the difference between a correct carrying out of that method and a mere unsuccessful attempt to carry it out. Second, methods themselves are subject to

normative appraisal: Some are good to use in some situations, others not so much. It's this second way that methods are involved with normativity that will matter most for my purposes here. This second way is multifarious, for there are many dimensions along with a method can be evaluated as good or bad.

Let me give you the central idea in broad strokes before getting down to the details of the theory. Causation, counterfactuals, and nomic necessity each appear to be intimately related to the normative appraisal of methods. For example, consider the counterfactual:

If you were to drop that glass on this stone floor, it would break.

Now consider the normative claim:

Under the present circumstances, dropping that glass on this stone floor is a excellent method for breaking it [the glass].

Most likely, you would be willing to assert the normative claim iff you were willing to assert the counterfactual. In fact, in many cases (and perhaps in most of the ones occurring outside of philosophy seminar-rooms), the main point of asserting the counterfactual might well be to convey the gist of the normative claim.

If you are a conspiracy theorist in the mold of Oliver Stone, then you might think the following counterfactual is true:

(1)   If Lee Harvey Oswald had not shot JFK, then someone else would have.

You might also think the following evaluative claim is true:

Stopping Oswald from pulling the trigger would not have been a particularly good way of saving JFK's life.

3

By contrast, if you believe the Warren Commission when it says that JFK's assassination was the act of a crazed gunman acting alone, then you probably believe the contrary evaluative claim:

> Stopping Oswald from pulling the trigger would have been an effective way of saving JFK's life

and you probably deny the counterfactual (1). Either way, whatever it is that motivates you take the counterfactual stand you do is probably exactly the same as whatever it is that motivates you to take the evaluative stand you do.

The connection is if anything more obvious for causation. If regular exercise causes good heart health, then exercising regularly is a good method of promoting heart health. If, by contrast, the practice of buying a new pair of fancy expensive running shoes every year is merely positively correlated with good heart health, but not a cause of it, then buying the shoes every year is not a good strategy for helping your heart.

Maybe it's less obvious that there is a connection between nomic necessity and norms about methods, but there is one nonetheless. Good methods can be reliable to different degrees, but at one end of the spectrum are methods that are perfectly reliable, in the sense that, as things happen, every single time they are carried out correctly, their ends are realized. If $\forall x(Fx \supset Gx)$, then there's a method for making sure that the item under my bed is G, by making sure that it is F – and ex hypothesi, every time this method ever gets correctly carried out, the thing in the box turns out to be G. But clearly, most such methods are no good at all. There must be countless universal regularities that are true only as a matter of complete accident[1]; if $\forall x(Fx \supset Gx)$ is one of these, then the method I just described is not a commendable method. Anyone who used it and then found a G under my bed would be indebted to dumb luck for his apparent success.

---

[1] See e.g. Armstrong 1983 for numerous arguments for this claim.

By contrast, suppose some method's reliability is underwritten by a law of natre. For example, suppose that we live in a universe in which it is a law that momentum is always conserved. Then here is a very nice method for making sure that system S has total momentum M at time t: First, at an earlier time t', make sure that S has total momentum M, and then make sure that S remains isolated from its environment until time t. Now, that might not be the easiest trick to pull off. But it seems undeniable that, if only you could do it, it would be an excellent method for making sure that S has M at t. When it works, it isn't just a fluke. The method is not merely uniformly successful, but it is also a commendable method, one you would be wise to use. And the regularity that assures us that this method is unfailingly successful is just the law of conservation of momentum.

And so in general: Every method that works perfectly every time it is correctly implemented corresponds to some true universal regularity that expresses its unfailing success – we might cal this the method's *success condition*. The methods like this that are not just uniformly successful as a matter of contingency are the ones whose success conditions are nomically necessary. Thus, the problem of distinguishing the truly commendable methods from the ones that just happen to work every time, and the more familiar problem of distinguishing laws of nature from accidental uniformities, are really the same problem in different guises.[2]

Thus, the modally-inflected notions of causation, counterfactuals, and nomic necessity seem to be correlated in an important way with norms concerning the quality of methods. I reckon most philosophers would be inclined to think that this is because which methods are the good ones is determined in part by the counterfactuals, the causal relations, and the laws. But I'm going to try out the hypothesis that the order of explanation really goes the other way: The counterfactuals, the causal relations, and the laws and nomic necessities are all grounded in the norms about which methods are the good ones.

This hypothesis might strike you as a non-starter; it might seem obvious that the modal notions of law, cause, and counterfactual are more fundamental than the

---

[2] Later on, I'll return to the point just made, and sharpen it up somewhat.

normative notion of a good method for finding something out or getting something done. But is that really so obvious? One thing that might tempt you into thinking so is the fact (and I agree that it is a fact) that when we want to figure out what would be a good method to use for some purpose, we typically start by thinking about what we know about the causal relations and the laws. But al this shows is that in particular cases, the order of discovery often runs from the modal to the normative, rather than vice versa. But the epistemic order of discovery frequently comes apart from the metaphysical order of dependence. Perhaps that happens here.

“Still,” you might ask, “isn’t it much more plausible that the modal things ground the normative ones than the other way round? Normative facts and features are thought to be ‘queer,’ as Mackie put it – they are not the sort of thing to serve as a basic building block for a respectable sort of universe.” Well, the struggle to reduce the normative without remainder to the non-normative still goes on in divers regions of philosophy, and the end of the tunnel does not really appear to be in sight. What seems plausible to me is that we are stuck with al sorts of normative things whether we like it or not, and only a pious (or is it impious?) hope gives us any reason to suspect that they can be reduced away. Moreover, this doesn’t mean that we have to accept normative facts as “basic building bocks of the universe” – there are on-realist and non-reductive options for understanding the normative, such as projectivism and quasi-realism (see e.g. Blackburn (1986)).

And there are considerable advantages to be gained by starting out with normative notions and trying to construct the natural modalities out of them. For this approach makes it easier to account for the links between modal and normative notions we’ve just been glancing at. E.g., most of us would agree that if regular exercise is a cause of good heart health, whereas annual expensive shoe-buying is only spuriously correlated with it, then it would be a good idea to take up exercise for the sake of your heart but a bad idea to take up shoe-buying with the same end in mind. But why is this? If causal relations are simply reflections of the norms governing which methods we should use, the matter is perfectly transparent: What it is to be a cause of some effect is simply (roughly) to be a good kind of means to

use when you want that effect – so you don't want to be using any methods that aren't causes of the things you want.

By contrast, suppose that causal relations are a non-normative matter. Then why should we prefer exercising to buying shoes? What I want is for my heart to be in good shape. So when I'm choosing among the options I believe to be open to me, while heart health is on my mind, it makes sense that I should look for an option such that if I pick it, I'll likely have good heart health. According to the received wisdom, however (which – let me quickly note – I agree with), that's not quite right: What I should look for is an option such that if I pick it, then I'll likely thereby *cause* my heart to be healthy. But look, I don't care whether I cause my heart to be in good shape or not – I just want it to be in good shape. It's not as if I fetishize causing things myself.[3] Those of us who are not causal fetishists and who care about our hearts should be willing to settle for having healthy hearts, whether we caused them to be that way or not. And yet, we aren't – most of us take the information that annual shoe-buying is, though correlated with, not a cause of heart health to settle it that the shoe-buying strategy is a lousy one to pursue. If causation is the kind of normative notion I'm suggesting it is, then this is almost a tautology – if not, then it's a frustrating puzzle.

This is at best a motivational argument, though. What really matters is what kind of theory you can get if you pursue the normativist strategy – how illuminating, how unifying, how plausible it looks once the whole theory is on the table. So, let's get on to the detailed work.

One more short comment before we do, though. In what follows, I will be offering a number of analyses of some things in terms of other things, and making some claims about the metaphysical order of dependence. I will also be taking some things as primitives. Of course, there is a lot of meta-philosophical discussion going on these days about just what philosophers are doing, or ought to be doing, when they engage in this sort of thing. In this paper, I'll have nothing to say about this. This will be an exercise in going ahead and doing the philosophy first, and leaving

---

[3] Maybe some people do; I read somewhere recently that Heidegger felt guilty about living in a universe that he did not create himself. But that's just weird.

the meta-philosophy for later. The owl of Minerva flies only at dusk, or something like that.

2. A Short Discourse on Methods

2.1. Some Varieties of Methods

The first task is to get clear about what I mean by a "method," and about what's going on when we evaluate methods. As I mentioned at the beginning, I have two sorts of methods in mind: Practical methods, which are methods for getting thigns done, and observational methods, which are methods for finding things out that involve input from sensory experience. An observational method needn't be *purely* observational – making several telescopic observations and then feeding their results into a calculation counts as an observational method, in my sense. But the experiential input cannot be idle; it must make a difference to the result. The method of doing an arithmetic problem in your head as you casually look out the window doesn't count as an observational method for finding a sum. Actually, I think we can probably reduce observational methods to a special case of practical methods, and get by with only one sort of method in our theory. But for purposes of the present paper, I doubt it's worth the effort.

By a "method" (of either kind), I mean roughly a type of process that a suitably-placed agent might exploit for purposes of achieving something, but which might also occur naturally without the involvement of any agent. For example, a nice method for shattering a glass bottle is to see to it that a rock with great momentum collides with it. When Suzy and Billy throw rocks at bottles, they exploit this method to further their destructive ventures, but of course, collisions of rocks with bottles can happen independently of anyone's schemes. Terminological stipulation: I'll say that a method is *correctly carried out* or *successfully instantiated* or just *instantiated* whenever the process in question occurs, whether it does so owing to the intentions of some agent or not.

8

Any token act of an agent – indeed, any token event – will presumably be an implementation of many different methods, of finer and coarser grains. For example, when Suzy throws a rock at a bottle and breaks it, her act successfully implements the method of seeing to it that a glass bottle breaks by means of seeing to it that it has a rough collision with a rock, but this act also successfully implements the method of seeing to it that (say) a Pepsi bottle breaks by means of seeing to it that it has a run-in with a chunk of obsidian, and so on. No doubt, she only means to be carrying out one (or anyhow, a small number) of these methods, and she carries out more fine-grained ones accidentally, so to speak. We might say that she *deliberately* carries out only those methods that she had specifically in mind, and that she carries out the other ones *incidentally*.

I assume that a method can be carried out collectively by a group of agents, either with or without deliberate coordination. Whenever this happens, there also exists a method for the same end which only one of the agents performs. For example: Beaker carefully calibrates a scale, which Dr. Honeydew then uses to weigh a rabbit. Collectively, the two of them have carried out a certain observational method for finding out the weight of a rabbit, which we might roughly characterize as the calibrate-the-scale-then-put-the-rabbit-on-it method. There's another method in the neighborhood here, which Dr. Honeydew has carried out all by himself – the method of getting-a-scale-your-lab-assistant-tells-you-is-properly-calibrated-and-putting-the-rabbit-on-it. Both methods have been carried out there. Beaker and Honeydew collectively carried out one of them, while Honeydew carried out the other alone – even though in an intuitive sense Honeydew only did one thing. I will call methods like the one Honeydew carried out alone a *deferential* method, since it involves relying on another agent to correctly carry put some other method. (There are deferential observational methods, too, like the method of finding out the mass of Mars by Googling it.)

There is a special case that I need to mention briefly. Sometimes, one method will be related to another in a ay that makes it natural to say that it "piggybacks" on it. For example, here is one method for measuring the mass (as opposed to the weight) of a rabbit. First, put the rabbit on an ordinary, well-calibrated spring scale,

9

and get a reading for its weight. Then drop a cannonball from the top of a tower and time its descent. Carry out the relevant calculation and determine the local gravitational attraction from the time of the cannonball's descent. Divide your result for the rabbit's weight by your result for the local gravitational acceleration. The quotient you get is the result of your rabbit-mass measurement. Let's call that method *the full scale method*.

Here's an easier method for measuring the same thing: Put the rabbit on the scale and read off its weight. Divide the result by 9.8 meters per second per second. The quotient is your result. Let's call this one *the shortcut scale method*.

They're both good methods for measuring the mass of a rabbit. And the shortcut method takes a lot less time and trouble, and it might even be more accurate most of the time. But there's at least one dimension of goodness along which the full scale method is clearly better: It leaves less to chance, so to speak, and this is reflected in the fact that it is evidently more counterfactually robust than the shortcut method – had the size and/or mass of the Earth been different, then the local gravitational acceleration would have been different, and the shortcut method would not have still been reliable. But the full scale method still would have been.

In general, what makes a method a *piggyback* method is that it can be generated from a second method by introducing a certain sort of shortcut. Sometimes a correct carrying out of one method necessarily involves carrying out another method along the way – in this case, we can say that the second method is a *submethod* of the first. (A practical method can have either a practical or an observational method as a submethod; an observational method can have either a practical or an observational method as a submethod.) If you take one method that contains a submethod, and you modify it by replacing the submethod with some other step that, as a matter of fact, always leads to the same result as the submethod – or at least, is known to have a high objective probability of yielding the same result – then you have a method that *piggybacks* on the original.

2.2. Metaphysics of Methods

There are lots of metaphysical questions you might ask about methods. As I said, I think of a method as a kind of process that a suitably-placed agent might exploit for some end. But please don't take this as a technical definition or a metaphysical analysis. For official theoretical purposes, I take the notion of a method as a primitive, and I will make a few basic assumptions about methods.

First of all, a method can be correctly carried out – and in every fully-specified possible situation, there is a fact of the matter about whether it was correctly carried out or not. Thus, it has a well-defined set of possible correct-carryings-out. If you like, you can think of methods as individuated by their sets of possible correct-carryings-out.

But not just any old set of possible token processes gets to count as picking out a method. For, I assume, a method has to be something that can be intelligibly seen as a case of going-on-in-the-same-way. As Wittgenstein taught us, we can hardly get by without assuming that we have a basic notion of going-on-in-the-same-way, and it is very hard to see how to reduce it to anything more basic. (In this direction lies the rule-following regress…) Someone might invoke a primitive notion of a natural classification here, and say that the methods can be identified with the sets of possible token processes that belong to some natural classification. I have no real objection to that. Speaking for myself, though, I can't see that anything is further illuminated by such talk, and I think I have a much better primitive grasp of what "going on in the same way" is supposed to be than I do on what a natural classification" is supposed to be. So, I'll stick with taking methods as a primitive, adding by way (hopefully) of at least a little clarification that a method has to be a way of going on in the same way.

A particular method can be specified by a set of parameters. There are many possible formalisms you might adopt for this purpose; here is mine:

A particular practical method is specified by an ordered triple of propositions <M, E, C>, where the *means* in this method is seeing to it that M, the *end* is seeing to it that E, and the *enabling condition* for the method is C. So, <M, E,

C> is the method of seeing to it that E by means of seeing to it that C under conditions C.

A particular observational method is specified by an ordered 4-tuple <Q, P, K, R> where Q and P are magnitudes or quantitative properties, K is a proposition, and R is some kind of positive correlation relation – perhaps identity, perhaps some mathematical function, perhaps some kind of statistical relation. Q is called the *object variable* of the method; it is the quantity getting ostensibly measured. P is the *pointer variable* of the method; it is the variable whose final value is the result of the putative measurement; K is the enabling condition for the measurement, which specifies everything that must be done in order for the method to have been correctly carried out (how the equipment is constructed, etc.); R is the correlation relation that ostensibly holds between Q and P whenever the measurement has been correctly carried out, i.e. when K is true.[4]

2.3. Evaluating Methods

Methods can be evaluated along more than one dimension. Evidently, a necessary condition for a method to be worthy of consideration at all is that it be generally successful. Let's say that a practical method <M, E, C> is *generally successful* just in case $Pr(E|M.C) > Pr(E|C)$, were Pr is some suitable objective probability measure (such as the global relative frequencies). In other words, under conditions C, E happens more often when M happens than otherwise; the means boosts the probability of the end. There are at least two dimensions along which we might measure the extent to which a method is generally successful: along one dimension, the method is more generally successful the greater the probability of E given M&C is; along the other, the method is more generally successful to the extent that, given C, M boosts the probability of E. Let's say that an observational method

---

[4] It might not really, though; what I have just defined is an observational method, not a *good* observational method.

<Q, P, K, R> is *generally successful* just in case it is true that throughout space and time, whenever K obtains, R holds between Q and P.

This is not an unrealistic idealization. It does not mean that such a method "always works perfectly," much less that things always go well when we try to use it, or even that things have always gone well when we justifiably believe that we have used carried out the method correctly. Recall that K can be a probabilistic or statistical correlation relation rather than a strict correlation. Since there are many kinds and degrees of correlation, there are many dimensions along which observational methods can be evaluated for general successfulness. Moreover, most of the methods e actually employ are such that our sincerely trying to carry them out correctly does not suffice for their having been carried out correctly. All that is required for an observational method to be generally successful is that among cases where C really is true, R holds between Q and P.

General successfulness, to whatever degree, s not al there is to being a good method. For one thing, a method can be generally successful just as a matter of accidental fact – its success can be a great fluke. Its general success might also be based on what we usually call a "spurious correlation," like in the shoe-buying method discussed in section 1. So there have to be other dimensions of goodness for methods as well. Perhaps the obvious thing to say now is that a method is really commendable only if its means causes its end, or if its end depends counterfactually on its means, or something along those lines. I don't want to say anything like that, though, because my aim is to show how to analyze the natural modalities in terms of methods and their dimensions of goodness; it would be viciously circular of me to take for granted a causal or counterfactual dimension of goodness.

Instead, I'm going to assume that there is another dimension of goodness for methods that is irreducibly normative. I hereby appropriate the word "effectiveness" as a technical term for this dimension. A practical method that has at least some positive degree of effectiveness is, I believe, what is generally known as an "effective strategy" (see Cartwright 1979). The maximally effective observational methods are what I will call "complete measurement methods" – and I will try to

13

show by examples below that that this matches pretty well with the way we ordinarily use the term "measurement." Now for the definitions:

> A practical methods <M, E, C> is *maximally effective* =def an agent A's having seen to it that M in conditions C can be a good reason to attribute *unshared responsibility* (or *credit* or *blame*) to A for the outcome E.

> An observational method <Q, P, K, R> is *maximally effective* (and so, a complete measurement method) =def an agent A's having seen to it that C and ascertained the value of P can be a good reason to attribute *unshared responsibility* (or *credit*) to A for discovering the result that the Q = [the value of P].

> By contrast, a method (of either type) is *not effective to any positive degree at all* – no matter how generally successful it is – if an agent's having carried it out correctly cannot be a good reason for assigning her any degree of responsibility for the outcome. For example, the mere fact that I have bought a fancy new pair of running shoes every year of my adult life in itself is no reason at all for giving me any share of responsibility for the healthy state of my heart.

That gives us the extremes – maximal effectiveness, and total ineffectiveness. A method has some intermediate degree of effectiveness if it can provide a good reason to attribute some shared or derivative responsibility for an outcome. The degree is higher the less widely the responsibility is shared and the less derivative it is.

For example: One observational method for finding out the mass of Saturn is to make a certain set of telescopic sightings, record the results, and plug them into a certain set of calculations – see Newton's *Principia* for the details of what sightings you'll need to make and what calculation you'll need to do. That, I take it it, would deserve to be called a method of *measuring* the mass of Saturn. If you carried that

method out all by yourself, you would arguably deserve a full share of credit for empirically finding out the mass of Saturn. (Not that you're the first person ever to find out the mass of Saturn. Lots of other people deserve unshared credit for this too. The sense of "unshared" in play here is not the sense in which it is logically impossible for two different people both to deserve unshared credit for the same thing.) Another obsevraitonal method for finding out the mss of Saturn is to Google it. That is an extremely good method for finding out the mass of Saturn: It is generally successful to an extremely high degree. But nobody would call it a way of *measuring* the mass of Saturn. A person who used this method might deserve *some* credit for finding out the mass of Saturn – after all, we do say things like "My brilliant seven-year-old daughter found out the mass of Saturn on her own, by Googling it." But her responsibility in this case is presumably both shared and derivative. For her success derives from the earlier success of astronomers who actually measured the thing.

So an observational method can be an effective method even if it isn't a complete measurement method (i.e. even if it isn't *maximally* effective). Moreover, as this case illustrates, the dimension of effectiveness is orthogonal to that of general successfulness: The method of finding the mass of Saturn by Googling it might have a much greater success rate than that of making a fresh astronomical measurement. Still, the latter is more effective than the former – the latter is a genuine measurement method, whereas the former is not. Th same thing goes for practical methods.

There are other dimensions of goodness for methods as well. Methods can be more or less feasible, they can be more or less inexpensive, they can have fewer or more undesirable side-effects, they can be more or less stylish, or funny. I won't be concerned with any of those other dimensions of goodness here. For the purposes of the natural modalities, I think it turns out that the only dimensions of goodness of methods that matter are the ones I have called general successfulness and effectiveness.

2.3. The Epistemology of Method-Evaluation

How are we supposed to find out how effective some method is? As I hinted above in section 1, we often do this by starting out with what we already know about causes, laws, and so forth, and draw inferences from these about which methods are the good ones to use in various circumstances, guided more or less by the principle that, roughly speaking, the effective methods are the ones that are not merely generally successful, but also counterfactually reliable – that would still have been generally successful even if things had been different in a wide variety of ways. The epistemic order of discovery needn't coincide with the metaphysical order of dependence.

But I cannot just say that and leave it there. For one thing, how does our knowledge about methods and causes and so forth ever get started in the first place? For another, we evidently learn new things about causes, laws and counterfactuals as science progresses – this can't all have been implicit in what we knew before, so where does this new knowledge come from? Moreover, on the view I'm proposing, talk abut laws and causes and counterfactuals is "at bottom" a way of conveying information about the normative statuses of methods, which strongly suggests that we ought to be able in principle to get along without ever using causal, nomic, or counterfactual vocabulary and just talk directly about methods and their qualities. (In fact, I suspect that we could do that, in principle, but it would be outrageously inconvenient.) So there ought to be some principled story about how we go about finding out how good methods are along my two key dimensions that doesn't appeal to our knowledge about natural modalities.

So I need an epistemology of method-evaluation. I'll just provide a sketch of one here. I think we find out which methods are generally successful simply by using induction and statistical inference. Whatever the right epistemology of induction and statistical inference is, just plug that in here. With effectiveness, things are hairier.

We start out having always already thought that some methods are effective. Crying out is an effective method for getting parental attention; putting an object in your mouth is an effective method for finding out about some of its most interesting

16

properties. I assume that if we find ourselves already believing that some method is effective to some degree, then we are prima facie justified in continuing to believe it so.

Similarly, I think that whenever we discover, e.g. via induction, that some method is generally successful to some appreciable degree, that gives us a prima facie justification for believing that the method is maximally effective.

But there are some exceptions to these rules, as well as some conditions under which our prima facie justifications get defeated.

First off, whenever we learn that some method is not generally successful to any degree at all, we lose al justification for thinking that it is effective to any degree. Effectiveness to any degree requires general successfulness to some degree.

Second off, there is an exception for deferential methods: When an observational method involves taking someone's testimony, in any form (including by using Google), it cannot be a maximally effective method. Similarly, whenever a practical method involves getting another agent to do something for you, it cannot be a maximally effective method either. (Though it should be noted, that you and your teammates might collectively carry out a method that is indeed maximally effective.) Being *deferential* in these ways knocks a method down in the effectiveness hierarchy by as many rungs as there are links in whatever chains of deferrals it involves. (Though you and your team members may defer to one another, collectively, as a team, you need not defer to anyone.)

Third off: Prima facie justification for believing a strategy to be effective is defeated whenever we have evidence that there is another effective strategy for the same end, such that the second strategy screens off the statistical correlation of the first one with the end.[5]

---

[5] Alternatively, one might want to say that in order to get the prima facie justification, one first has to do due diligence in checking to see whether or not this defeat condition is satisfied.

More specifically: Suppose the means in the first strategy is M1 and the means in the second strategy is M2. Let E be the common end of both strategies. Then although:

$$Pr(E|M1) > Pr(E|{\sim}M1)$$

nevertheless:

$$Pr(E|M1 \ \& \ M2) = Pr(E|{\sim}M1 \ and \ M2)$$

and:

$$Pr(E|M1 \ \& \ {\sim}M2) = Pr(E|{\sim}M1 \ \& \ {\sim}M2)$$

in other words: If you hold fixed whether means M2 is employed, the question of whether M1 is employed makes no difference to the probability of E. Under this condition, strategy M2 screens off M1's influence on E. So, if we are justified in believing that M2 is an effective strategy for E, this defeats whatever prima facie justification we may have had for believing that M1 is an effective strategy for E.

This is why we are not justified in believing that buying an expensive pair of running shoes every year is an effective strategy for avoiding heart disease: Among people who exercise regularly, presumably, the probability of heart disease is the same among those who do and those who don't buy an expensive pair of running shoes every year, and similarly among those who don't exercise regularly. And we are justified in believing that regular exercise is an effective strategy for avoiding heart disease, because we are justified on inductive grounds in believing that it raises the probability of avoiding heart disease, and we are not justified in believing in any other effective strategy that screens it off in the way that it screens off the strategy of buying new shoes every year.

Of course, there's nothing new in this requirement that the influence of an effective strategy not be screened off, but I think the way I'm implementing the requirement here is unusual. A more familiar implementation of the idea is found for example in Nancy Cartwright's (1979) account of the relation between type-level causation and laws of association: She holds that an event-type C causes and event-type E just in case C raises the probability of E within every otherwise causally homogeneous situation. In other words, if you hold fixed every factor other than C that is causally relevant to whether E occurs, then E is more probable given that C

occurs than that it doesn't.[6] This account creates an epistemological problem: We can never be sure that we know all the factors that are causally relevant to whether E occurs, so we can never be sure whether we are looking at the statistics within the right populations when we assess causal claims. Of course, there are ways of dealing with this problem. But the way I am implementing the no-screening-off condition neatly sidesteps this issue: The thought is that in order for us to be justified in believing that a strategy is justified, we must not be justified in believing in the effectiveness of another strategy that screens off the first one. This doesn't require us to know, or even be justified in believing that we have anything like an exhaustive list of the causally relevant factors on hand.

Fourth off: our prima facie justification for thinking that a generally successful method is maximally defective is defeated whenever that method piggybacks (in the sense explained back in section 2.1) on another method that we are justified in believing to be effective to any degree. The piggyback method must be lower on the effectiveness hierarchy than the method it piggybacks on.

Fifth off, and finally: In general, when our justificaton for believing a method to be effective to one degree is defeated, but not in a way that requires us to regard as completely ineffective, then we are prima facie justified in believing it to be effective to the next lower degree, and we remain so until new defeaters come in.

This gives us a sketch of an epistemology for method evaluation. But why should believe this is the right epistemology? My answer is as follows. I think the best account of the notion of an effective method is a conceptual-role semantics. A conceptual-role semantics needs two parts: the input-part, which specifies the conditions under which you are justified in applying that concept, and the output-part which specifies what else you're committed to once you have applied that concept. In the case of effective methods, the output-part is the normative/evaluative role of the concept, that I talked about earlier. So what about the input-part? I suggest that the input part is constituted by a certain set of epistemic rules for making judgments about effectiveness of methods. I think you

---

[6] Cartwright 1979, p. 423.

can formulate those rules without mentioning causation – and I think I just made a good start at it.

In a way, my defense of the above epistemology is very simple: The epistemological rules above (and perhaps some others in the same family) are constitutive of the concept of an effective strategy, and that's just all there is to it. But that answer is a little disappointing: What we really want to know is why the epistemological input rules I've proposed for the concept of an effective strategy are tied together with the normative/evaluative output rules I've described, to make one concept? Why not put those output rules together with some totally different set of input rules, and get a concept that plays the same role in practical reasoning as our concept of an effective strategy does but is governed by totally different evidential norms?

It seems to me that the kind of answer that question calls for is an explanation of why it makes sense for creatures like us to use the concept of effective strategy that I've described, rather than some alternative one. Given the kinds of interests we have, the kinds of capacities we have, the kinds of lives we lead and the kinds of lives we want to lead, it is sensible for us to conduct our practical deliberation and our evaluation of each other's activities with the aid of a concept of effective strategy that is governed by these rules rather than some other set of rules. That is, it is predictable that things will go better for us if we use these rules then if we use any alternative available set of rules we might have used instead. I'm optimistic that an argument to that effect can be given, but I have to admit I don't have all the details yet, and I'm not going to go any further into that question here.

3. The Primitives

With these preliminaries out of the way, we can get into the theory itself. First, the primitives:

As mentioned above, I take *practical methods* and *epistemic methods* as primitive notions (though I suspect that we can reduce one to the other if we really

want to). Though I don't offer an analysis of either, I have made a number of assumptions about what they are like, laid out in section 2.1 and 2.2 above.

A particular method is specified by specifying the value of a few parameters – these parameters take propositions, magnitudes, and correlation relations as values. The correlation relations are just logico-mathematical relations, and I assume that enough of mathematics is there for me to appeal to these relations without taking on board any substantive primitives. The magnitudes are just ordinary quantitative properties, like mass, weight, velocity, pH and so forth, and the propositions that appear in specifications of methods are what just propositions about ordinary natural things, their properties and relations. So I assume that all those sorts of things exist. But I do not make any further metaphysical assumptions about their natures. In particular, I don't assume that the properties and magnitudes are universals, or tropes, or resemblance-classes or whatever. I also make no assumptions about whether they have essences, or about their modal profiles. Unlike Humeans, I do not start out assuming that they are quiddities, susceptible of unlimited recombination, and unlike essentialists, I do not assume that they have dispositional essences. The question of their modal profile is one that I think should eventually be settled by the normativist theory itself (though I won't get that far in this paper), not built in with the primitives.

4. Causation

The normativist theory I recommend takes a page from the agency, i.e. manipulatonist or interventionist, approach to causation, which was pioneered by von Wright (1971) and has been lately developed to great sophistication by Woodward (2003) and by Menzies and Price (1993) (in very different ways). The basic idea behind this approach is that what it is for C to cause E is for C to be a good way that somebody could bring about E.

There are four important objections that any verson of the agency approach is going to have to come to terms with:

First, *the circularity objection*: Agency, manipulation, and birnigng-about are themselves causal notion, so you can't appeal to it in an analysis of causation, on pain of circularity.

Second *the inaccessibility objection*: We think that small inhomogeneities in the early universe caused the formation of galaxies. But obviously no agent is able to bring about the formation of galaxies by means of making there be small inhomogeneities in the early universe. More generally, not all causes can be manipulated by agents, and this looks like a problem for the agency approach to causation, because it isn't easy to see how it's going to capture the fact that those non-manipulateable causes are causes.

Third, *the anthropocentrism objection*: Causation seems to be a fundamental feature of the universe; agency seems to be a very special and local feature of some animals. So it seems inappropriate to analyze causation in terms of agency. To do so is to imply that a very special feature of human beings is somehow built into the structure of the cosmos. And that seems weird.

Finally, the *token-level problem*: Even if there is something to the agency approach to causation, it appears that at best, it can give us only an account of type-level causation; things are going to get very hard when we turn to token-level causation. To see why, consider the beloved example of Billy and Suzy, two juvenile delinquents throwing rocks at bottles in an alley. Throwing a rock at a bottle is a good way for an agent to get a bottle to break, and perhaps this is what grounds the type-level causal relation between rock-throwing and bottle-breaking. But in the case at hand, Suzy throws her rock a little harder than Billy throws his, so Suzy's rock gets there first and shatters the bottle, leaving only a cloud of glassy debris for Billy's rock to fly through. Billy and Suzy both did things that in general are god means to getting bottles to break, but only Suzy caused this bottle to break. This is of course the problem of overdetermination, which famously ramifies in

complicated ways, spawning lots of counterexamples to lots of theories. It isn't easy to see how the agency approach can solve this problem.

I suspect that a resolutely normativist version of the agency theory of causation, resting on the primitives I introduced in section 3, can solve all four problems.

My primitives give me the notion of an *effective strategy*, which for me is just a practical method that is both generally successful and effective to any positive degree at all. I think this notion coincides with the familiar notion of an effective strategy, which came to prominence in the discussion of causation mainly due to Nancy Cartwright's paper "Causal Laws and Effective Strategies."[7] An effective strategy is a good way of bringing about some specified outcome. For example, getting regular exercise is an effective strategy for avoiding heart disease. (Not a foolproof strategy, but nevertheless an effective one.) On Cartwright's view, effective strategies have a lot to do with causation, and it is impossible to understand what makes an effective strategy effective solely in terms of regularities and statistical correlations – not even if those regularities and correlations are underwritten by the laws of nature.

For example (again), there is a very strong positive statistical correlation between being in the habit of buying a new, expensive pair of running shoes every year, on the one hand, and avoiding heart disease, on the other. But this doesn't show that buying a new pair of expensive running shoes every year is a good strategy for avoiding heart disease. That correlation exists only because people tend to shop for shoes like that only if they are serious runners, and serious runners get plenty of exercise, which helps them avoid heart disease. But just buying the shoes every year isn't going to help your heart.

On Cartwright's view – which has become pretty standard – we have to understand effective strategies in terms of causal relations, and we cannot understand them just in terms of laws, regularities, and statistical correlations. So,

---

[7] Cartwright 1979.

causal eliminativism and causal reductionism are both off the table. We have to take causation seriously as a basic feature of the world.

I agree with almost everything Cartwright said there, but I think she got one thing upside down: It isn't that the notion of an effective strategy depends on the notion of causation; it's the other way around: causation should be analyzed in terms of effective strategies.

So the basic idea – which we will need to refine a bit – is this:

TYPE-CAUSATION: **C** is a type-level cause of **E** in conditions **K** just in case seeing to it that an event of type **C** occurs is an effective strategy for seeing to it that an event of type **E** occurs in conditions **K**.

TOKEN-CAUSATION A: C is a cause of E iff: C happens, E happens, C and E are wholly distinct events, and some condition K prevails such that C is an instance of an effective strategy for bringing about E type in conditions K.

You might think this proposal is going to fall prey to the Circularity Objection, You might say, "This is cheating! You can't analyze causation in terms of effective strategies! Isn't that idea just a non-starter? Isn't it obvious that causation is the more fundamental thing here, and that effective strategies need to be analyzed in terms of it? Isn't it obvious that the concept of an effective strategy is itself a causal concept, and the property of being an effective strategy is a causal property, so that any analysis of causation in terms of effective strategies would have to be viciously circular?"

And the answer is: No, that's not obvious. Why should we think that the notion of an effective strategy presupposes the notion of causation? One reason might be that in order to figure out which strategies are the effective ones, we usually have to rely on what we already know about what causes what. That's true, but it doesn't imply that causation can't be analyzed in terms of effective strategies. Once again, the order of epistemic discovery in particular cases doesn't have to coincide with the order of analysis.

Anyhow, what seems to be essential to effectiveness of strategies is the role that it plays in our reasoning about practical action. In particular, it plays a normative role, and it plays an evaluative role. The normative role comes up when we are deliberating about what to do: Calling a strategy "effective" is a way of marking it as one that it may be worth considering when we are interested in bringing about the relevant end. The evaluative role comes up when we are evaluating the activities of other agents: Which outcomes it makes sense to charge to a particular agent's account, so to speak, depends on what things they have done that are effective strategies for those outcomes. For example, to a first approximation: I can't hold Robert responsible for the mess on the floor unless I hold that he did something – either on purpose or inadvertently – that I take to be an effective strategy for producing a mess on the floor. So the concept of an effective strategy plays these roles, and it seems clear that any concept that did play these two roles would be functioning as the concept of an effective strategy – it would *just be* the concept of an effective strategy. In short, an effective strategy just is a *good* way of getting something done. This is a normative/evaluative concept. It's not fundamentally a causal concept.

You might argue that even so, there's still a way in which the concept of an effective strategy presupposes causation: An effective strategy is always an effective strategy *for* something, namely for *bringing about some outcome.* But *bringing about* an outcome is, of course, a causal notion.

My reply is that you could look at it that way, but you don't have to. An effective strategy is a strategy for getting a certain thing done. What is *getting a thing done*? Well, whenever you get something done, you do cause something, for sure – but that only shows that getting things done is conceptually related to causation – it doesn't tell us which way the direction of dependence goes. What it means to say that an agent A got a certain outcome O done is, I suggest, to credit the outcome to the agent's account, so to speak. It's a normative thing as well, not fundamentally a causal one. So much for the Circularity Objection.

The Inaccessibility Objection is no problem. *We* cannot use the strategy of seeding the early universe with little lumps in order to bring about galaxy

formation. But al that shows is that this is not a *feasible* method of making galaxies –
it doesn't show that it isn't an *effective* one. And indeed, it obviously is.

Let's be clear about what it means for something to be an effective strategy:
An effective strategy is a strategy such that *if* an agent *were* to bring about the
means (perhaps *per impossibile*), that could be a good way for them to go about
bringing about the effect. (If you could make two tectonic plates press against each
other for a few hundred million years, that would be a great way to make a
mountain range. If only you could do it. The fact that you can't do it, and neither can
anybody else, only shows that it is not a feasible strategy, not that it isn't an effective
one.)

So, characterizing something as an effective strategy does not entail that it is
within the capacities of any existing agent to carry out. That's why the kind of
account I'm proposing can get around the inaccessibility objection. What's more,
characterizing something as an effective strategy does not entail that there exist any
agents at all. So there can be effective strategies even in a world without agents in it.
And in that way, the kind of account I'm proposing can overcome the
Anthropocentrism Objection as well.

Now, you might wonder: Why do we have a concept of an effective strategy
that can apply to strategies we don't have the ability to use? Isn't that wasteful?
Couldn't we get by with a more modest concept, that only applies to strategies we
can actually use?

Answer: It's a good thing that we have a concept of effective strategy that has
a variable extension, so we can add new things to its extension as we learn more
about our environment. It is also good for the way this extending happens not to be
willy-nilly, but to conform to certain rules – to keep our practical reasoning running
along useful lines. But there isn't any good way to write a rule that restricts the
things that can count as effective strategies to things we can actually carry out. This
is because we don't yet know what things we're going to be able to actually carry out
in the future – simply because we don't yet know which ends we're going to find
effective strategies for. So, it made good sense for mother nature to endow us with a
concept of effective strategy that allows us to add more things to its extension

26

without any restriction on the kinds of things that can be means. What's more, in particular, it's helpful to be able to find out that some means-end pair is an effective strategy before we know of any effective strategy for bringing about that means. (That can help us decide whether it would be worthwhile to spend time trying to figure out how to bring about the means.) For these reasons, it makes sense for our concept to be such that means-ends pairs can get to be counted as effective strategies even when we don't have any way of using the means available to us.

This leaves us with only the Token-Level Problem. It's certainly possible for tow agents each to carry out, separately, two effective strategies for the same end, but for only one of them to count as causing a realization of that end. This is the familiar problem of excluding cases of pre-emption and trumping. One obvious strategy for ruling those out as non-cases of causation is simply to exclude them by brute force. We can do this by modifying TOKEN-CAUSATION-A to yield TOKEN-CAUSATION:

TOKEN-CAUSATION: C is a cause of E iff: C happens, E happens, C and E are wholly distinct events, and some condition K prevails such that C is an instance of an *untrumped* and *unpreempted* effective strategy for bringing about E type in conditions K.

It might look as if "untrumped" and "unpreempted" are weasel-words, but in fact, you can analyze each of them in terms of effective strategies too:

> C* *trumps* C as a cause of E iff: C*, C and E all happen; some condition K prevails in which C is an effective strategy for E; some condition K* prevails in which C* is an effective strategy for E; and a modified version of C* is an effective strategy for a modified version of E under conditions (K, K* & C), but it is not the case that a modified version of C is an effective strategy for a modified version of E under conditions (K, K* and C*)

More colloquially: C* trumps C as a cause of E iff a tweaked version of C* would be an instance of an effective strategy for a tweaked version of E even under conditions where C still happens, but a tweaked version of C would not be an instance of an effective strategy for a tweaked version of E under conditions where C* still happens.

So for example[8]: The sergeant and the major both shout orders the soldiers to advance simultaneously, and they do. Both the sergeant's shout and the major's are effective strategies for getting the soldiers to advance. But having the major shout "retreat" is an effective strategy for getting them to do something else, under conditions that include the sergeant's still shouting "advance"; by contrast, having the sergeant shout "retreat" under conditions where the major is still shouting "advance" is not an effective strategy for getting them to do something else. So, the major's shout trumps the sergeant's shout. So the major's shout is a cause of the soldiers' advance, but the sergeant's shout is not.

Now for preemption:

C* *preempts* C as a cause of E iff: C*, C and E all happen; some condition K prevails in which C is an effective strategy for E; some condition K* prevails in which C* is an effective strategy for E; E+ is a more fine-grained (i.e., more fragile) version of E, C*+ is a more fine-grained version of E, and there are prevailing conditions under which C*+ is an effective strategy for E+, whereas there is no more fine-grained version of C which is, under prevailing conditions, an effective strategy for E+.

More colloquially: C* preempts C as a cause of E iff C*'s happening in just the way it did, but not C's happening in just the way it did, was an instance of an effective strategy for E's happening in just the way it did.

---

[8] This example is attributed to Bas van Fraassen by Lewis 2000.

So for example: Suzy and Billy each throw a rock at a bottle; Suzy's bottle gets there first, and it smashes the bottle. Both throws are effective strategies for breaking the bottle. But a throw with the precise physical characteristics of Suzy's is also an effective strategy for getting the bottle to break in the precise manner that it did, whereas a throw with the precise physical characteristics of Billy's is not. So Suzy's throw preempts Billy's her throw is a cause of the bottle's breaking, and Billy's is not.

Thus, we can reduce even singular causation without remainder to the normative notion of an effective strategy. No doubt, there are other hard cases to worry about. But this seems a very promising start.

5. Counterfactuals

I suggest that one of the central uses – perhaps *the* central use – of counterfactuals in non-philosophical discourse is to comment on the goodness or badness of methods. In section 1, I already considered a few illustrative examples. More are easy to multiply, and even infamous cases of weird philosophical counterfactuals seem to fit the pattern: Back during the Korean Conflict, what would have been a good way to get some catapults to be used as weapons in combat there? Well, it might not be very feasible, but if you could just get a Roman general like Julius Caesar in command there…   What would have been a good way, back during the Korean Conflict, of getting the atomic bomb used? Well, it might not be that feasible, but if you could just get a totally ruthless general like Julius Caesar in command there… I bet you know the rest of this story.

The rough intuitive idea that I want to start with is this: The counterfactual A $\square\!\!\rightarrow$ C is true just in case someone here in these actual circumstances with an overriding desire that C be the case would be well-advised to see to it that A. That seems roughly equivalent to saying that:

29

COUNTERFACTUAL-A: Where A is false, A $\square\!\!\rightarrow$ C is true iff there is a good method (local terminological stipulation: a *good* method is one that is both generally successful to some positive degree and effective to some positive degree) M such that M's success condition SC_M entails A $\supset$ C.

(Recall that a method's *success condition* is the true generalization – be it strict or statistical – that expresses its general successfulness.)

That's not quite right, though. For one thing, since A is supposed to be false, there might be some halfway decent method whose actual reliability depends on A's being false. That shouldn't' be enough to make A $\square\!\!\rightarrow$ C true; but according to COUNTERFACTUAL-A, it is. So let's modify:

COUNTERFACTUAL-B: Where A is false, A $\square\!\!\rightarrow$ C is true iff there is a good method M such that M's success condition SC_M entails A $\supset$ C, but SC_M does not entail ~A.

There's still a worry, though. Suppose that the condition on the right-hand side of the biconditional in COUNTERFACTUAL-B is true, but there is also a good method M' such that $SC_{M'}$ entails A $\supset$ ~C without entailing ~A, and suppose that M' is at least as effective as M itself is. In that case, our rough intuitive thought – that the counterfactual is true just in case an agent with an overriding desire that C would be well-advised to see to it that A – would clearly be false. And obviously, we should reject the counterfactual A $\square\!\!\rightarrow$ C.

But can cases like that really arise? Here is an example. Let A be the proposition:

A: The Earth's mass and radius were lately modified in such a way that the local gravitational acceleration is no longer 9.8 meters per second per second, and you now measure the mass of a rabbit using the shortcut scale method[9], which you carry out correctly.

_____

[9] See above, circa page 10.

and now let C be the proposition:

C: You get the right result for the mass of the rabbit.

The success condition for the shortcut scale method is basically the generalization that says that that method delivers an accurate result whenever it is carried out correctly. So, COUNTERFACTUAL-B implies that A $\square\!\!\rightarrow$ C is true. But we know that this counterfactual is in fact false: If you followed the instructions for the shortcut scale method, after the local gravitational acceleration had changed, you would get an incorrect result for the rabbit's mass – you would get the rabbit's weight divided by 9.8, whereas the correct result is the rabbit's weight divided by g, where g is not equal to 9.8. We know this because we now that the *really proper* way to measure the mass of a rabbit would be to use the full scale method. The success condition for the full scale method tells us that the full scale method gives an accurate result whenever it is correctly carried out – which implies that if the local gravitational acceleration is not 9.8 meters per second per second, then the shortcut method will give the wrong answer. Thus, the full scale method in effect assures us that under the stipulated conditions, carrying out the shortcut method would be a great way to get the wrong answer.

Of course, by the same token, the success condition of the shortcut method in effect assures us that under the stipulated conditions, the full scale method would be a great method for getting the wrong answer. This isn't an even standoff, though: The full scale method wins, because it is the more effective method – because the shortcut method piggybacks on it. Thus, this is indeed the kind of example that raises trouble for COUNTERFACTUAL-B.

But the trouble is eased by switching to:

COUNTERFACTUAL-C: Where A is false, A $\square\!\!\rightarrow$ C is true iff there is a good method M such that (i) M's success condition SC_M entails A $\supset$ C, but SC_M does not entail ~A, and

(ii) for every good method M' such that $SC_{M'}$ entails $A \supset \sim C$, but $SC_{M'}$ does not entail $\sim A$, M is a more effective method than M'.

The second clause allows the full scale method to block the shortcut method from underwriting the false counterfactual $A \mathbin{\square\!\!\rightarrow} C$, but it stops the shortcut method from blocking the full scale method from underwriting the true counterfactual $A \mathbin{\square\!\!\rightarrow} \sim C$, precisely because the full scale method is a more effective method than the shortcut method. And in general: When a counterfactual supposition pits two good methods against each other, the more effective one prevails.[10]

Just one more complication, now. It might be that seeing to it that A is a good method for seeing to it that C, , but that it wouldn't *guarantee* C, but only make it probable. I cases like tis e often balk at sating, "Had it been the case that A, then it would have been the case that C." Instead, we often prefer to say, "Had it been the case that A, then it *would likely* have been the case that C." (Or, "*would very likely …*" of "*would almost certainly …*", or "*would with probability p …*")[11] Taking account of this complication, here is my final proposal about counterfactuals:

COUNTERFACTUAL: Where A is false, $A \mathbin{\square\!\!\rightarrow}$ (likely, etc.)C is true iff there is a good method M such that (i) M's success condition SC_M entails $A \supset$ (likely, etc.)C, but SC_M does not entail $\sim A$, and (ii) for every good method M' such that $SC_{M'}$ entails $A \supset$ $\sim$(likely, etc.)C, but $SC_{M'}$ does not entail $\sim A$, M is a more effective method than M'.

---

[10] Notice that what matters here is which method is more effective, nt which one is more generally successful. Even if it were the case that, for some strange reason, random errors happen more frequently with the full scale method tan with the shortcut method, so that the shortcut method was generally successful to a greater degree, it will still be obvious that if the local gravitational acceleration were different, the full scale method would still be reliable whereas the shortcut method would not.
[11] Hajek (ms) discusses cases like this at great length, and gives a rather different treatment of them from the one I give here.

This analysis codifies the intuitive considerations we started out with, and it evidently gives the right verdicts on a very wide range of counterfactuals – though there is still more testing to be done.

6. Laws and Nomic Necessity

Marc Lange has argued in a number of places (e.g. 2009) that the kind of necessity enjoyed by the laws of nature is aptly characterized in terms of their relation to counterfactuals. In particular, Lange argues, a species or grade of necessity worthy of the name *necessity* should be such that when we characterize it precisely, and define the correlative sense of "possible" using the standard equivalence ◊P iff ~□~P, then we find that the following truism is vindicated:

> What is necessary is what would still have been true, not matter what possibility had been actualized

or more formally:

$$\forall P(\Box P \text{ iff } \forall Q(\Diamond Q \to Q \,\Box\!\!\to P))$$

It is not too hard to see that a given set of truths comprise the necessities, in a sense of "necessity that passes this test, just in case the set is *stable* in a sense defined by Lange:[12]

> A set of propositions G is *stable* =def for any p that is consistent with G, and any q that is in Q, $p \,\Box\!\!\to q$.

---

[12] I have modified and simplified Lange's formulation by suppressing complexities that make no difference here.)

Lange proposes the view that the laws of nature are the members of the largest non-trivially[13] stable set and nomic necessity consists in being a member of this set. Counterfactuals Lange takes as a primitive; it is, on his view, because the counterfactuals lie as they do that the laws lie as they do.

I agree with all of this. It's worth noticing though that on Lange's account, there is no particular reason why there should be any non-trivial stable sets of truths in our world, or why we should expect there to be. If there is one, then that seems to be a kind of brute fact about our world – one of amazing scope and complexity, and which just happens to accommodate many of our intuitions about the laws of nature. The theory of counterfactuals I offered in the preceding section, however, both predicts that, and explains why, there exists a stable set. And along the way, it offers us a nice hypothesis about what lawhood really is.

Consider the set of all generally successful methods that are maximally effective. (Among observational methods, these include the complete measurement procedures, but they do not include the methods that piggyback on them. Among practical methods, it includes the maximally effective strategies, but not the deferential ones.) Since these methods are all generally successful, they all have true general success conditions. Now consider the set of all those success conditions; call this set **L**. This set, I claim, must be stable, given the semantics for counterfactuals I have proposed.

Here's the argument: Suppose that C is in **L**, and that A is consistent with **L**. Then C is the success condition of some maximally effective method M*; thus C = SC_{M*}. Now recall:

COUNTERFACTUAL: Where A is false, A $\square\!\!\rightarrow$ (likely, etc.)C is true iff there is a good method M such that (i) M's success condition SC_M entails A $\supset$ (likely, etc.)C, but SC_M does not entail ~A, and (ii) for every good method M' such that $SC_{M'}$ entails A $\supset$ ~(likely, etc.)C, but $SC_{M'}$ does not entail ~A, M is a more effective method than M'.

---

[13] The set of all truths is trivially stable, because there is no counterfactual supposition consistent with it.

This implies that given what we have stipulated about A, C, and $\mathbf{L}$, $A \,\square\!\!\rightarrow C$ is true. To see why, Let $M = M^*$. Then $C = SC_{M^*} = SC_M$. Hence SC_M entails C, ergo it entails $A \supset C$. But SC_M does not entail $\sim A$, since SC_M belongs to $\mathbf{L}$ and A is consistent with $\mathbf{L}$. Now suppose that M' is some method such that $SC_{M'}$ entails $A \supset \sim C$. Then $SC_{M'}$ entails $A \supset \sim SC_M$. So, $\{SC_M, SC_{M'}, A\}$ is an inconsistent set. Therefore $SC_{M'}$ must not belong to $\mathbf{L}$, since SC_M does belong to $\mathbf{L}$ and A is consistent with $\mathbf{L}$. Therefore M' is not a maximally effective method. But M is, so M is a more effective method than M', as COUNTERFACTUALS requires. Hence $A \,\square\!\!\rightarrow C$ is true. But A could have been any false proposition consistent with $\mathbf{L}$, and C could have been any member of $\mathbf{L}$. Therefore, $\mathbf{L}$ is a stable set, QED. This result follows from the semantics for counterfactuals defended above, and the assumption that there exists a set of maximally effective methods; those are enough to show that there must be a stable set.

$\mathbf{L}$, recall, is the set of all the success conditions of the maximally effective methods. By naming it $\mathbf{L}$, I foreshadowed my next proposal, which is that $\mathbf{L}$ is the set of laws, and lawhood just is the property of being the success condition of a maximally effective method.[14]

This hypothesis appears to be extensionally adequate: If some method were not such that its general success was guaranteed by the laws of nature, then its success would be, well, contingent: Anyone who relied on it and achieved success would be beholden to a lucky contingency. It wouldn't be right to take a method like that to be the basis for an attribution of primary, unshared responsibility for the success. I.e., it wouldn't be right to call it a maximally effective method. So, the success condition of any maximally effective method must be nomically necessary. And conversely: If we knew that some methods success condition was nomically necessary, then we would know that carrying out that method correctly was a nomically sufficient condition for success. What could be a more effective method

---

[14] This theory of lawhood is very similar, but not identical, to one I defended in my 2008 and to a similar one I defended in my 2012. The main difference here is that practical methods are included alongside observational ones.

than one like that? So, the nomically necessary regularities seem to give us blueprints for building maximally effective methods, of which they serve as the success conditions.

Therefore, we have good reason to think that the nomically necessary regularities are the very same truths as the success conditions of the maximally effective methods. Of course, to say *this* is only to assert an extensional equivalence. My proposal is the stronger one that this equivalence holds because the lawhood of the laws is grounded in the maximal effectiveness of the maximally effective methods. This completes the delivery of what was promised: A plausible, unified account of causation, counterfactuals, and nomic necessity that presents the, all as grounded in normative facts about which methods are the good ones to use.

WORKS CITED:

Armstrong, D. M. 1983: *What is a Law of Nature?* Cambridge University Press.

Blackburn, Simon 1986: "Morals and Modals," in Graham MacDonald (ed.) *Fact, Science and Morality*, Blackwell.

Brandom, Robert 2008: *Between Saying and Doing: Toward an Analytic Pragmatism*, Oxford University Press.

Cartwright, Nancy 1979: "Causal Laws and Effective Strategies," *Nous* 13(4): 419-437.

Hajek, Alan (ms): "Most Counterfactuals are False."

Lange, Marc 2009: *Laws and Lawmakers.* Oxford University Press.

Lewis, David 1973: *Counterfactuals.* Harvard University Press.

Lewis, David 2000: "Causation as Influence," *Journal of Philosophy* 97(4): 182-197.

Menzies, Peter and Price, Huw 1993: "Causation as a Secondary Quality," *British Journal for the Philosophy of Science* 44(2): 182-203.

Roberts, John T. 2008, *The Law-Governed Universe*, Oxford University Press.

Roberts, John T. 2012, "Measurement, Laws, and Counterfactuals," in Mumford and Tugby (eds.) *Metaphysics and Science*, Oxford University Press.

Ryle, Gilbert 1950/1971: "'If', 'So', and 'Because'" in *Collected Papers,* Volume 2. London: Hutchison.

Sellars, Wilfrid 1954: "Some Reflections on Language Games," *Philosophy of Science* 21: 204-228.

Thomasson, Amie 2007: *Philosophical Topics,* 35/1&2, 2007: 135-160.

von Wright, Georg 1971: *Explanation and Understanding.* Cornell University Press.

Woodward, James 2003: *Making Things Happen*, Oxford University Press.