Weighted Explanations in History

Robert Northcott

Philosophy of the Social Sciences 2008; 38; 76
DOI: 10.1177/0048393107311045

Philosophy of the Social Sciences
Volume 38 Number 1
March 2008 76-96

10.1177/0048393107311045
http://pos.sagepub.com
hosted at
http://online.sagepub.com

Weighted Explanations in History

Robert Northcott
University of Missouri, St. Louis

Weighted explanations, whereby some causes are deemed more important than others, are ubiquitous in historical studies. Drawing from influential recent work on causation, I develop a definition of causal-explanatory strength. This makes clear exactly which aspects of explanatory weighting are subjective and which objective. It also sheds new light on several traditional issues, showing for instance that: underlying causes need not be more important than proximate ones; several different causes can each be responsible for most of an effect; small causes need not be less important than big ones; and non-additive interactive effects between causes present no particular difficulty.

Keywords: *causation; explanation; history; interaction; proximate; underlying*

"Every historical argument revolves around the question of the priority of causes."

—E. H. Carr (1962, 117)

1. Introduction

Was the fall of Rome due more to barbarian invasions or to internal decay? What was the main cause of the American Civil War? Was Saddam's regime or international sanctions more responsible for the deaths of Iraqi children? In such cases, all agree on the putative causes; dispute focuses only on which of them is the more explanatorily important. *Weighted explanations*, whereby some causes are deemed more or less important than others, are ubiquitous in historical studies and indeed everyday life. But it turns out that furnishing a good account of them is a surprisingly delicate task, and one so far treated either unsatisfactorily or not at all in the explanation and

Author's Note: I would like to thank Raymond Martin, Jim Bohman, Paul Roth, Steve Turner, and Alison Wylie for useful help, and the latter four also for organizing the 2007 Roundtable. I am grateful too for the feedback from audiences at the Roundtable, and at the St Louis philosophy of science reading group.

Received 6 September 2006

philosophy of history literatures. As a result, it is still unclear what exactly a historian is claiming when offering a weighted explanation, and also unclear exactly what kinds of evidence are relevant to assessing such claims. "[This issue] is a fair field full of inexplicit and unconscious assumptions . . . tactfully concealed from the public in the interest of avoiding intellectual scandal."[1]

It proves very useful to make our concepts in this area explicit. Drawing from influential recent work on causation and causal explanation, in this article I develop a new definition of causal-explanatory strength. Among other things, this yields a principled way to incorporate pragmatic aspects of explanation, via formal relativization to contrast classes in both the cause and effect slots. It also makes clear exactly which aspects of explanatory weighting are subjective and which objective.

We may not always be able to assign exact values to the quantities in the definition, relying instead on plausible estimates. But still we can make progress in understanding causal strength and explanatory weight even without settling the epistemological issues. There is value in purely conceptual work. First, we thereby clarify just how the justification of weighted explanations is epistemologically demanding. Second, and the main focus of this article, we also become newly able to disentangle several mistaken debates.

In part 2 of the article, I cover the necessary technical preliminaries. Then after those greens, so to speak, the dessert: in part 3, I apply these preliminaries to show why many widespread claims and assumptions regarding weighted explanations are, surprisingly, either false or confused. For instance, it turns out that: underlying causes are no more important than proximate ones; several different causes can simultaneously each be responsible for most of an effect; non-additive interactive effects between causes are no particular hindrance to apportioning explanatory responsibility; and small causes may be equally as explanatorily important as big ones.

2. Technical Preliminaries

2.1 A Definition of Causal Strength

Define the *strength* or *importance* of a cause C with respect to an effect E to be:

$$E(C_1 \,\&\, W) - E(C_0 \,\&\, W) \qquad (1)$$

1. Historian J. H. Hexter, quoted (Martin 1989, 53).

where C_1 is the actual level of C, C_0 is a baseline counterfactual level of C, and W is background conditions. In the simplest case, C_0 will just be the absence[2] of C.

Formula (1) is quite intuitive. We are interested in the quantity of effect for which C is responsible, and this is just the level of effect with C compared to the level with some alternative input. For example, the causal strength (CS) of kicking a ball might be yielded by the ball's acceleration with the kick compared to its acceleration without that kick. A key aspect of (1) is that it captures a controlled-experiment sensibility. We want to compare the level of effect with and without the cause while *keeping all else equal*. For instance, it would be no use comparing the acceleration of a ball with and without a kick if simultaneously a gust of wind had blown up, because obviously the calculation would now yield only the combined impacts of the two changes. For this reason, in (1) the background conditions W must be constant across the two terms.[3]

It is obvious that the strength of any given cause depends on the effect in which we are interested. (A kick may be a powerful cause of accelerating a ball, but not of accelerating a chemical reaction.) In our notation, that is, a C's CS is E-specific. But notice also an important further corollary of the above discussion—namely that, even once E is specified, still C's CS has no unique value. Instead, for any given C-E pair, the associated CS will depend also on two further things:

1. *Background conditions W*. For example, striking a match will cause a light if W includes sufficient oxygen in the atmosphere but not otherwise.
2. *Contrast class* C_0. Often, as with temperature or genotype, the "absence" of a cause may make little sense. Rather, in such cases we are interested in the impact of a cause relative to some specific non-zero alternative. This also often tracks normal usage. For instance, in sports the causal impact of bringing on a bench player will depend both on that player *and* on which player was being replaced. Thus in its right-hand term, (1) cites the general formulation C_0 rather than "absence of C" or some such.

---

2. Although I omit discussion of the exact definition of events, no commitment need be implied here or elsewhere to "negative events." More generally, throughout I shall switch between understanding C_1, E_1 etc as values of the C and E variables, and understanding them as specific events. Nothing important turns on this formal looseness though.

3. Strictly speaking, in fact the background conditions *do* vary across the formula, because as well as impacting E, in general the switch from C_0 to C_1 will also change W as well. But for our purposes we may ignore that technical wrinkle.

So overall, the notion of CS is necessarily a highly relativized and context-specific one, a fact that will prove significant. In particular, for any given cause-effect type pair, there will be *many* associated token causal strengths—a new one for every change of W or C_0. No C-E pair has any univocal CS *simpliciter*.

2.2 Counterfactuals

$E(C_0 \& W)$, the right-hand term in (1), is a *counterfactual*—we are interested in what the level of effect *would* have been, given C_0 and W. How can we ever know the values of such counterfactuals? Because, strictly speaking, W is never quite exactly the same from moment to moment, epistemologically the best we can ever do is to find data from as good a re-creation as possible of the relevant conditions. (This is precisely the rationale, for instance, governing the design of controlled experiments.) In this respect, (1) serves as a *normative ideal*, telling us what hypothetical quantity is relevant, which then serves as a guide to how best to estimate that quantity from *actual* data. Of course, only some actual data, namely that best approximating controlled constant-W conditions, will be appropriate.[4]

An immediate and important corollary is that: to the extent that the relevant counterfactual cannot be evaluated, so a CS cannot be either. For example, if a ball's acceleration without a kick were for some reason unknowable, then so would be that kick's CS. The pressing issue here, of course, is that whereas counterfactuals may be relatively easy to evaluate in an experimental setting, often matters are much more difficult in history. For instance, how many Iraqi child deaths would there have been in the absence of sanctions? We might analyze child death rates from the pre-sanctions era, from similar countries unburdened by sanctions, and so on. If it is thought that we simply cannot know, then it becomes correspondingly impossible to quantify those sanctions' CS. (1) merely guides us how to try.[5]

---

4. For example, in the 1849 London cholera epidemic there was a strong inverse correlation between infection rate and altitude. But still the CS of altitude was not high. This is because dirty water (which *did* have high CS) also inverse-correlated with altitude. The right-hand term in altitude's CS formula, which should have been for a constant level of dirty water, thus could not be estimated from the raw actual data. Generally, going counterfactual addresses in this way the problem of confounds.

5. I do not endorse any particular semantics for counterfactuals here, as for our purposes the salient locus of philosophical dispute lies elsewhere.

2.3 Subjective and Objective

In light of the above, are assignations of CS objective? It is crucial to be clear on the senses in which they aren't—and on those in which they are. Begin with a couple of *subjective* aspects. First, as we saw, any CS depends on choice of contrast class C_0, and that in turn is interest-relative. (Of course, once *given* a particular interest, choice of C_0 may thereafter be objective.)

Second, the choice of causal *vocabulary*, i.e. of E and C variables, is also partly interest-relative, and moreover perhaps more likely to be controversial in history than in physical science.[6] On the other hand, any given CS is obviously specific to a C-E pair anyway, so it is not really any further problem if for each new choice of C and E we also have a new CS. It follows that if two historians are using different vocabularies their CS claims may well not be in competition with each other, depending on how inter-translatable those vocabularies are. Meanwhile, I need not be committed to there being any canonical categorization or individuation of historical causes. (See section 3.4 for discussion in the context of an actual example.)

Moving to the objective side, in any particular case the actual background conditions W (as opposed to their description) are presumably objective, as of course is the actual level of effect, i.e. $E(C_1 \& W)$. Summarizing: once given a vocabulary (i.e. C, E) and a context (i.e. C_0, W), a CS is thereafter objective. Therefore the notion of CS is *not* hopelessly mired in arbitrariness; rather, it is merely a *relational* concept. Relativity does not imply subjectivity. CS is perfectly objective and non-arbitrary, no less than other relational concepts such as compass bearing or relative velocity.

What of the counterfactual $E(C_0 \& W)$? Whether one thinks its value is objective will turn on one's views regarding counterfactuals generally. In accordance with the philosophical mainstream, I think a complete modal skepticism is implausible. In any case, the objectivity of counterfactuals is a presupposition of CS talk. Nevertheless, as noted earlier, even then still *epistemological* difficulty may remain, namely that it may be especially hard to evaluate $E(C_0 \& W)$ in the non-experimental situations typical of historical investigation. Indeed, arguably counterfactuals from history are notorious for exactly this reason. Notice though the distinction between a counterfactual being merely hard to evaluate, on one hand, and a counterfactual having no objective value even in principle, on the other. The former does not imply the latter. In particular, difficulties with the epistemology of

6. "Vocabulary" here is meant to encompass choice of ontology too.

counterfactuals do not at all threaten the *metaphysical* objectivity of (1) as a definition of CS.

Like all counterfactuals, those in (1) may be formulated ambiguously. In such cases, then so is the CS correspondingly ambiguous. Indeed, arguably a constraint on choice of vocabulary is precisely that the associated counterfactuals are sufficiently well defined to have objective meaning. But history is no different in this respect from social or natural science generally.

Whether attributions of CS are objective has long been a central concern of the philosophy of history literature. I think that (1) encompasses the insights of both sides. On one hand, choice of salient contrast class C_0 may often reflect normative concerns. For instance, was the salient alternative to sanctions against Iraq some lesser degree of sanctions, no sanctions at all, favored trade status, extensive development aid, or military action? Normative concerns may also often impinge on choice of E_0 (section 2.6 below). On the other hand, as noted, all this is perfectly consistent with CS still being objective once these other parameters have been specified.

2.4 Probabilities

When E is a quantitative variable, matters are unproblematic. But if an effect is dichotomous—i.e., an event that either happens or doesn't—then we need to interpret $E(C_1 \& W)$ as the *probability* of the effect occurring. More generally, in indeterministic cases we may interpret $E(C_1 \& W)$ as the expected value of the effect variable. Formally, this is straightforward enough. But the probabilities involved are objective single-case ones, and philosophically those are notoriously problematic.

A classic difficulty is the so-called problem of the reference class. For example, my calculated chance of developing cancer will vary greatly depending simply on how I classify myself. As a thin person with a healthy lifestyle I may have a low chance; as a male with some family history of cancer I may have a high chance. The point is that neither choice of reference class seems any more objectively compelling than the other, and thus my objective chance too seems correspondingly arbitrary. An exactly analogous problem arises in historical cases, as Turner (2007) emphasizes. The issue applies equally to both the left-hand and right-hand terms in (1). And choice of reference class is still underdetermined even once given choice of C and E.

In response, some interpretations of single-case probability, such as the propensity interpretation or Lewis's best-system view, claim to avoid the

reference class problem, although of course these views have other difficulties of their own. In any case, even if avoided metaphysically, still the reference class problem may resurface epistemologically. We may only be able to estimate the propensity of Franz Ferdinand's assassination to lead to European war, for instance, by reference to other situations in history we regard as usefully analogous—and that latter choice of situations, the criticism runs, is precisely what lacks objective justification.

Nevertheless, for better or worse, I think the objectivity of single-case probabilities is a presupposition of CS talk. In mitigation, the issue is a problem for everyone, and not just when constructing weighted explanations. That is, appeal to such probabilities is necessary right across much of science. Biologists, for instance, appeal to them all the time, yet we do not on that account give up on objective biological inquiry. Moreover, the closely related issue of establishing best rational credences occurs throughout social science and decision theory. It is not obvious that history is especially badly off—especially as choice of reference class seems only seldom to be the locus of actual historical dispute. In practice, historians take it to be perfectly possible to get some handle on the estimation of these probabilities via indirect methods. In any case, finally, the points to be made in part 3 below do not seem hostage to whichever particular view of single-case probability we eventually endorse.

2.5 Relation to Previous Literature

Formula (1) reflects the common emphasis in the analytical literature on causation's *difference-making* aspect: a cause is something that makes a difference to its effect. Thus, naturally, the strength of a cause is how *much* difference it makes. Much philosophical debate has focused on which—if any—exact formulation of difference-making might be an acceptable *definition* of causation. Fortunately we need not worry about that here, as our concern is not with what the putative causes of an effect are but rather only with which is the strongest of a set of causes already agreed on by all. The existing literature specifically on the latter issue is relatively sparse. But the exact formulation of (1) is endorsed by the contemporary Bayes net and causal modeling literatures, and by experimental practice (Woodward 2003; Pearl 2000; Spirtes, Glymour, and Scheines 2000).[7] More generally,

7. Why not some other formulation besides the simple difference equation (1)? One reason is that only a difference equation captures the impact of interventions—i.e. the difference it made that we manipulated the world with C_1 rather than C_0. Such a manipulationist perspective is central to many contemporary theories of causation.

it is also consistent both with realism about causal powers, and with the mainstream literature on probabilistic causation. (For further philosophical details, see Sober, Wright, and Levine 1992; Northcott 2006; Sober 1988; Lewis 1973.)

Turning to the philosophy of history literature, the motivation behind (1) is similar to that behind several classical views, for instance those in the nineteenth century of Yule and Weber (Turner 1986). Historically, both von Kries and Weber, among others, recognized the context-specificity of CS, although less formally than does (1). Regarding CS's C_0-sensitivity, for instance, they instead took choice of description to set an appropriate equivalence class of not-C to guide the use of data. The counterfactual nature of (1)'s right-hand term was also not made explicit, CS sometimes seeming to be *defined* in terms of comparison with alternative actual data rather than such recourse to actual data being marked clearly as merely a method for evaluating the counterfactual quantity. Nevertheless, such modern-day nit-picking notwithstanding, the central intuitive idea was the same—namely, that CS is yielded by comparing the level of effect with the cause to that without it. (1) also shares a historical concern exclusively with physical causal responsibility, as opposed to moral or legal responsibility.

Notice though that (1) is far from being just an uninformative platitude. The benefits merely from being formally explicit will become apparent below. But in addition, another frequent strand in the philosophy of history literature is clearly rejected by (1) completely. I have in mind the linking of causal importance with causal necessity, which again dates back to the nineteenth century, and which has re-emerged in more sophisticated guise more recently, e.g. (Martin 1989)—see section 3.5 below. A separate tension arises with various statistical measures of CS used frequently throughout the social and behavioral sciences. These can be shown to be ill suited to their task, conflicting as they do with (1). Examples of such measures include the analysis of variance, and heritability (Northcott 2005b and 2006).

2.6 From Causal Strength to Explanatory Weight

While (1) is fine for CS independent of any explanatory context, recall that our original motivation was to make sense of weighted *explanations*. It turns out that, for this latter purpose, (1) must be augmented. To see why, consider an example now familiar in the literature: the bank-robber Willy Sutton was once asked "why did you rob a bank?" He replied: "because that's where the money is." Why is this amusing? Because Sutton's answer is appropriate to the question "why did you rob a bank *rather than some*

other store?", whereas the context naturally suggested instead the question "why did you rob a bank *rather than not rob anything*?" In other words, whether Sutton's answer is explanatory depends on the choice of contrast class on the *effect* side. (1) incorporates choice of contrast class only on the cause side—which is fine for an analysis of CS, but not for one of *explanatory* strength.

Generally, these days it is a familiar view that causal explanation has a pragmatic aspect, and going contrastive is the leading way to formulate this (Woodward 2003; Van Fraassen 1980; Achinstein 1983). Accordingly, contemporary theories of causal explanation are contrastive in both the cause *and* effect slots, taking the general form:

$$C_1\text{-rather-than-}C_0 \text{ explains } E_1\text{-rather-than-}E_0$$

If E is the effect variable, we may think of E_1 and E_0 as respectively the actual and contrast values of that variable. Intuitively, this captures the sensitivity of explanation to the precise specification of the explanandum.

How then to analyze the notion of explanatory *weight*? Most simply, if the change from C_0 to C_1 explains the change from E_0 to E_1, then it must be that:

1) E_1 occurs and C_1 occurs
2) E_0 *would* have occurred had C_0 occurred

If C_1 and E_1 are just the actual levels of cause and effect, condition 1 holds automatically. So a claim of full explanation in effect boils down to satisfying condition 2, i.e. to:

$$E(C_0 \,\&\, W) = E_0 \qquad (2)$$

The relativizations, and thus the objectivity, of this formula are the same as for (1), save for an additional relativization now also to choice of E_0, i.e. to precise specification of explanandum.

Formula (2) is extendable to probabilistic cases, for instance those with a dichotomous E. Generally, we can generate measures of the *degree* to which C has explained E (Northcott forthcoming). Such measures are novel and important. To my knowledge, the relativization to E_0 they encompass has been treated formally nowhere else. Sober, Wright, and Levine (1992) and Martin (1989) are two of the few even to emphasize the issue. For purposes of this article though, fortunately the simpler formula (1) will usually be sufficient. Thus, except where explicitly noted otherwise, I shall frame all examples so that E_0-sensitivity is not crucial, i.e. so that the E_0 implicitly

obvious from context will suffice. Hence for the most part we'll be able to analyze safely purely in terms of the less general—but more intuitive—causal strengths.

3. Applications

3.1 Limits versus Selections

A claim often heard, from Marxists and others, is that the deep structural factors underpinning history are the only truly important ones. For example, it has been argued that the threat available to capitalists of an investment strike, a threat that in the past has often been exercised, constrains the politically possible range of government policy (Przeworski and Wallerstein 1988). (The history of post-war Latin America may be one example.) This underlying structural constraint, the thought runs, is thus what really determines policy, not the mere surface matter of whichever choice some government then makes from the limited range of options still open to it.

But it is a mistake to privilege the deep structural factors as necessarily more important. Imagine that there are 30 different policies available—5 radical ones and 25 more conventional ones. Suppose that the capitalist threat structure limits those available to just the latter 25. Then it seems that a government official's particular selection, narrowing the field from 25 to 1, can be more explanatorily important than the threat structure, which narrowed the field merely from 30 to 25. To insist otherwise is merely to privilege one particular explanandum. Let us explicate this rather intuitive line of reasoning more formally.

Label: C_1 = the power of the threat structure, C_0 = no such threat power.

D_1 = the decision of the actual minister (who is as radical as the threats allow), D_0 = the decision of some alternative (but also radical) minister.

E_1 = actual non-radical government policy, E_0 = to be specified below.

Case 1) E_0 = one of the 5 radical policies.

Then, applying formula (2): $E(C_0 \,\&\, W) = E_0$, and $E(D_0 \,\&\, W) \neq E_0$. In words, without the threat structure there would have been a radical policy, whereas swapping ministers would have changed nothing. Thus it is the threat structure—not this particular choice of minister—that explains why we have the actual policy rather than a *radical* one.

Case 2) E_0 = the particular conventional policy from the 25 that would have been preferred by the alternative minister referred to in D_0.

Now, applying (2) again: $E(C_0 \,\&\, W) \neq E_0$, and $E(D_0 \,\&\, W) = E_0$. In words, swapping ministers would have led to the alternative conventional

policy, but removing the threat structure would not have. (Rather, the latter would have led to a *radical* policy.) Thus it is the particular choice of minister—and not the threat structure—that explains why we have the actual policy rather than the alternative *conventional* one.

In conclusion, there is no explanandum-independent sense in which the underlying threat structure has greater explanatory importance than the selection of the particular government official. Rather, explanatory importance is sensitive to choice of E_0 (as well as to choices of C_0 and W as usual). Although perhaps often forgotten or under-emphasized, this conclusion is hardly novel—indeed the structure of this very example is adapted from (Sober, Wright, and Levine 1992). But what *is* novel is the formal analysis, binding the conclusion to a well founded general theory of explanatory weighting.

3.2 Non-Additive Causal Interaction

Non-additive interaction occurs when the joint impact of several causes is not equal to the sum of their impacts individually. Is it possible or even meaningful to apportion explanatory responsibility in such cases? Many have thought not. Consider, for example: "On their own, none of mass unemployment, Hitler's charisma or widespread anti-Semitism made a Nazi government likely. Only combined were they so dangerous. So it is impossible to make sense of each factor's individual degree of responsibility—because each one's impact depended critically on its interaction with the others."

But it turns out that non-additive interaction in fact poses no special problem. Label for Germany in the 1920s and 30s: A_1 = occurrence of mass unemployment, B_1 = exercise of Hitler's charisma, C_1 = occurrence of widespread anti-Semitism. Let the effect variable E = the probability that a Nazi government came to power. For simplicity, assume that the salient contrast classes here are all absences: $A_0 = \sim A_1$, $B_0 = \sim B_1$, $C_0 = \sim C_1$. (As per footnote 2, no commitment is thereby implied to negative events.)

Recall that (1) tells us that any CS is yielded by "$E(C_1 \& W) - E(C_0 \& W)$." The exact numerical values will not be crucial to the point here, but for illustration's sake suppose that the probability of a Nazi government given all three of the factors was 0.8, but that this value drops off to varying degrees once one of the three factors is removed. Then (1) might yield these results:

> CS of mass unemployment = $E(A_1 \& W) - E(A_0 \& W)$ = pr(a Nazi government given mass unemployment) – pr(a Nazi government without mass unemployment) = 0.8 – 0.1 = 0.7, say

CS of Hitler's charisma = $E(B_1 \& W) - E(B_0 \& W)$ = pr(a Nazi government given Hitler's charisma) – pr(a Nazi government without Hitler's charisma) = 0.8 – 0.2 = 0.6, say

CS of anti-Semitism = $E(C_1 \& W) - E(C_0 \& W)$ = pr(a Nazi government given anti-Semitism) – pr(a Nazi government without anti-Semitism) = 0.8 – 0.3 = 0.5, say

So (for these figures and contrast classes), mass unemployment was more explanatorily important than anti-Semitism, for example, because it raised the probability of a Nazi government by a greater amount.

The real point is not the precise figures though. Rather, it is that these calculations are perfectly possible, or at least no less possible than other such calculations, *even though* quite consistently with them the example can easily be made one of non-additive interaction. Thus, reprising the key non-additive features from the original description:

"only combined were they so dangerous": $E(A_1 \& B_1 \& C_1 \& W) = 0.8$, the figure quoted already in the preceding calculations

"on their own, none made a Nazi government likely": easily represented by setting $E(A_0 \& B_0\, C_1 \& W)] = E(A_0 \& C_0\, B_1 \& W) = E(B_0 \& C_0 \& A_1 \& W) =$ 0.05, say

In words, the probability of a Nazi government is 0.05 when only one of A_1, B_1 or C_1 occurs, but is 0.8 (not 0.15) when all three occur, implying non-additive interaction.

I conclude that non-additive interaction is not a problem when assigning causal strengths. Intuitively, the key here is to remember that any CS is highly token, and the same cause may have many different strengths depending on W, i.e. on the levels of other factors. Here, for instance, mass unemployment has a very high CS in the presence of anti-Semitism and Hitler's charisma, but a very low one in their absence.

One consequence of the above is that in cases of non-additive interaction, different causes' strengths may sum to more than the total effect.[8] Here, each factor has an explanatory weight of more than half the total probability

8. At least, this is true of cases of *positive* non-additive interaction such as this example. Cases of overdetermination, by contrast, can be seen as ones of *negative* interaction, formally speaking. One such latter case, pressed by (Turner 2007), is the relative strengths of the different causes of the dropping of the Hiroshima A-bomb. In that example, each of the causes highlighted will likely have low CS.

of 1. But notice that no individual CS is greater than 1. Indeed, arguably we should *expect* the individual CS scores to sum to more than 1—that is the whole meaning of non-additivity! The take-home message is that *many different* causes may simultaneously have a high CS with respect to the same effect.

Finally, the analysis of this section also makes clear the solution to a traditional "ordering" worry (Turner 2007). That worry is that, in cases such as our example, the order in which we "subtract" causes determines their relative strengths. For example, if we first consider mass unemployment, this has a high CS, as we saw. But if, with unemployment then removed, we next consider anti-Semitism, the latter's CS will now be low. The reason is that, with unemployment already removed, the probability of a Nazi government becomes pretty low regardless of whether anti-Semitism is present. Yet if we had reversed the order and considered anti-Semitism first, and considered mass unemployment only with anti-Semitism already removed, the impact on CS scores would have been reversed. And clearly it is quite arbitrary which of the two causes we consider before the other.

But our scheme resolves this difficulty. All causes are treated symmetrically via the device of allowing multiple token CS scores for any given cause. Thus mass unemployment has a high CS if anti-Semitism is also present, but only a low CS otherwise—and *exactly the same is true*, mutatis mutandis, for anti-Semitism. Neither cause is privileged over the other, regardless of order of consideration.

3.3 Small Causes Can Be Important

"At Waterloo, the British fought the French to a standstill; the arrival of the Prussians was merely the straw that broke the camel's back. Clearly, the British contribution was therefore more important. Measurably asymmetric contributions imply asymmetric explanatory importance." But this turns out not to be true in general—even if we measure the British and Prussian contributions in an identical currency of number of soldiers and even if, further, the British number was indeed the greater. Rather, small causes can be as important as big ones. More fundamentally, when applied to a cause, the very designations "small" and "big" lack any absolute, context-independent meaning.

Label: C_1 = the fighting of the British soldiers, as measured by their number; D_1 = the fighting of the Prussian soldiers, as measured by their number; $C_0 = \sim C_1$ (i.e. as if zero British soldiers had been fighting), and likewise $D_0 = \sim D_1$.

For simplicity, denote Napoleon's actual defeat by E = 1, his non-defeat by E = 0. Also for simplicity, assume determinism (nothing important turns on this). Finally, assume—I believe historically correctly—that neither the British nor the Prussians would have won the battle alone.

Then, applying (1):

$$\text{CS of the British} = E(C_1 \,\&\, D_1 \,\&\, W) - E(\sim C_1 \,\&\, D_1 \,\&\, W) = 1 - 0 = 1$$

$$\text{CS of the Prussians} = E(C_1 \,\&\, D_1 \,\&\, W) - E(C_1 \,\&\, \sim D_1 \,\&\, W) = 1 - 0 = 1$$

That is, the British and Prussians have *equal* CS. This follows (in this deterministic case) simply from C_1 and D_1's individual insufficiency and joint sufficiency. Given this logical structure, the result holds *regardless* of the actual number of soldiers each country contributed. For instance, even if the British had 10 times as many men as the Prussians, still their CS was not any greater.

Recall again that any CS is highly context-specific. Thus with respect to many *other* explananda, say the number of French killed, of course the British had far greater explanatory importance than the Prussians precisely because of their far superior numbers. (Probably too our moral intuitions may be sensitive to such "type" considerations.) But with respect to *this* explanandum, we have no justification for awarding the British any more explanatory responsibility than the Prussians.

The temptation to assume that the Prussian force carried only a small CS is encouraged here by the *commensurability* of the numbers of British and Prussian soldiers. If there were more British than Prussian soldiers, the thought runs, then surely the British were a greater cause in some general sense? But this turns out to be a mistake, for the greater British force does not always represent a greater cause. Trivially, presumably the Prussians contributed more to the allied forces' German-language proficiency, for instance.

More generally, designating a cause 'small' is itself a claim regarding CS, but as we have seen any CS is context-specific. Thus, strictly speaking, there is no such thing as a cause that is "small" *simpliciter*; rather, a cause can only be a small or large cause *of something* (in context). Any cause, that is, is only large or small with respect to a specific effect in a specific context—including the British and Prussian forces in this example. Neither is a cause large or small in any absolute, explanandum-independent sense—the easy commensurability of number of soldiers merely misleads us into thinking so. And in this example, that tricks us into accepting incorrect asymmetric causal strengths.[9]

3.4 Underlying Versus Proximate Causes

"Underlying causes are more important than proximate ones." But this view, common in many sciences as well as history, is confused. Moreover, it remains equally confused even if we take the proximate cause to be itself an *effect* of the underlying one. To see why, consider an example. Label: B_1 = the CIA funding of the Bin Laden network in the 1980s, A_1 = Mohammed Atta forms his terrorist cell in Germany, E = probability of a major terrorist attack on the United States in the early years of the twenty-first century. For simplicity, set: $B_0 = \sim B_1$, $A_0 = \sim A_1$.

Applying formula (1):

> CS of the CIA funding: $E(B_1 \& W) - E(B_0 \& W)$ = pr(major attack in early 21C given the 1980s CIA funding to Bin Laden) – pr(major attack without that funding) = 0.7 – 0.1 = 0.6, say
>
> CS of the Atta cell: $E(A_1 \& W) - E(A_0 \& W)$ = pr(major attack in early 21C given the Atta cell in Germany) – pr(major attack without that cell) = 0.7 – 0.1 = 0.6, say[10]

Thus the two causes might plausibly be equally strong. Of course, these precise figures are somewhat arbitrary. But the point is not these particular figures, but rather how the calculation illustrates that there is no *general* reason to think that an underlying cause must be more important than a proximate one. The formula could easily have endorsed either of the two causes as the stronger.

It is also easy to see here how explanatory importance will be sensitive to the exact specification of E_0. If we are concerned to explain why there were attacks in 2001 rather than 2002, then likely A_1 will be more important, as Atta's particular cell was crucial to the attack's exact timing. If, by contrast, the explanandum is attacks in 2001 rather than some time after 2020, now Atta's cell is a mere detail. Presumably, given the strength of Bin Laden's network, some cell or other would have been ready to attack at some stage in the period, and thus now it will be B_1 that is explanatorily important.

The example also serves to illustrate other points from earlier. In particular, a different historian might have individuated causes differently, being concerned instead with the relative importance of, say, "American foreign

9. For fuller discussion of the issues in this section, see Northcott (2005a).
10. Notice that technically there is an interaction effect again here, which is why *both* causes might raise the probability by 0.6, i.e. by more than 0.5.

policy" and "Islamic militancy." These new causes' strengths could also have been calculated by applying (1). This new historian's claims about his or her own causes' CS values would then likely be largely independent of the other historian's CS values, and thus they might *both* be right. For instance, there need be no contradiction between the claim that Atta's cell was more important than the CIA funding of Bin Laden, and simultaneously the claim that American foreign policy was more important than Islamic militancy. There is no facile relativism here—each individual CS claim must be supported by arguments for specific evaluations of the relevant probabilities dictated by (1). Rather, the point is only to re-emphasize two earlier observations: first, that the availability of different choices of causal vocabulary does not somehow imply that objective CS evaluations are meaningless or impossible; and second, that there may, without incoherence, be many different major causes of the same event.

However, it seems that there does still remain one objection to my denial of privileged status to underlying causes. In particular, there is often one asymmetry that is undeniably context-independent, namely that the proximate cause is caused by the underlying one but not vice versa. For the sake of argument, concede that this claim is true in our example, for instance—i.e., that the CIA money was a (possibly distant) cause of the Atta cell but not vice versa. (Likewise, perhaps one of American foreign policy and Islamic militancy was a cause of the other but not vice versa.) Does this asymmetry not suggest that the underlying cause has greater importance in some absolute sense? This inference is not uncommon. The famous Roman historian A. H. M. Jones, for instance, argued in favor of the primacy of barbarian invasions as a cause of Rome's fall by noting that the chief rival suspect—internal decay—was itself in part an *effect* of those invasions (Martin 1989, 65f).

But this inference is faulty. In our example, even if the CIA money was indeed (ultimately) a cause of Atta forming his cell, and not vice versa, *still* it does not follow that the CIA money is any more explanatorily important. The asymmetric causal dependence simply does not enter the relevant CS calculations. In particular, it does not negate the fact that (on our figures) both factors raised the 9/11 attacks' probability equally. Intuitively, merely being an earlier link on a causal chain does not imply also being a stronger one. (See also Sober, Wright, and Levine [1992] on exactly this point.)

So how *is* the causal asymmetry between the underlying and proximate causes reflected formally in (1)? The answer is in asymmetric background conditions W in the two causes' respective CS calculations. In particular, W includes the (effects of the) Bin Laden money when calculating the CS of the Atta cell, but W does *not* include the Atta cell when calculating the CS of the

Bin Laden money. This is because W in (1) always refers to the background conditions at the time of the occurrence of C_1, which will include the effects only of all causes prior to C_1. (Strictly speaking, therefore, each formula should be quoting a different set of background conditions, "W_i" perhaps.)

Return, finally, to the importance of specifying precisely the explanandum, that is, in our notation, to E_0-dependence. For this purpose, consider another classical Marxist view, namely the endorsement of a material over ideological theory of history—the view that economic and technological changes are the primary drivers of world history, while ideological influences are merely symptoms of these underlying economic ones. What sense can we make of this claim?

Suppose for the sake of argument that while ideology does influence people's actions, it is indeed itself caused by economics rather than vice versa. For example, suppose that increased middle-class wealth leads to a stronger ideology for political rights. As a result, there is democratic reform. As just argued, still the two factors, economics and ideology, may have equal CS nevertheless: both were equally necessary links on the causal chain to reform. But suppose we are more precise about E_0, distinguishing between two alternative explananda:

1) E_0 = no democratic reform in the short run.

Now it may be ideology that has the higher CS: a change in ideology would have had a rapid impact on the prospects for reform, whereas the impact of a change in middle-class wealth would have taken longer to filter through (via an eventual impact on ideology). Thus the different ideology would imply this E_0, whereas the different economics would not. Formula (2) tells us that therefore only the former has high explanatory weight.

2) E_0 = no democratic reform in the *long* run.

Now the situation is reversed: an isolated change in ideology, unsupported by any change in middle-class wealth, would in time fizzle out – given our assumption that ideology is caused by economics. Thus the different ideology would *not* have implied this new E_0. An isolated change in middle-class wealth, by contrast, *would* have had a long-term impact on democratic reform, via its eventual impact on the intermediate causal link of ideology. Thus now it is economics, not ideology, which has the high explanatory weight.

I conclude that, depending on the detailed dynamics of the presumed causal structure linking economics, ideology and reform, the asymmetry

between economics and ideology may be reflected in asymmetric explanatory importances for particular explananda. But there is no sense in which economics, simply in virtue of being earlier in the causal chain, is therefore somehow more explanatorily important *simpliciter*. Rather, as is the general case, it will merely be more important with respect to some choices of E_0, and less with respect to others. The debate between the material and ideological theories of history, at least on this generalist construal of it, is therefore not substantive.

3.5 Causal Strength and Necessity

E. L. White remarked that if "mosquitoes were as necessary as the Christians [to the fall of the Roman empire, therefore] neither is paramount to the other" (Martin 1989, 54). The central thought here, as with other proposals in the philosophy of history literature, is to *identify* a factor's explanatory importance with its necessity for the explanandum. But such necessity is neither necessary nor sufficient for explanatory importance, as is easily demonstrated.[11]

Return first to the CIA/Bin Laden example, and interpret it as a fine-grained explanandum—that is, which factors raised the probability of the 9/11 attacks particularly, rather than some other attack early in this century? Suppose that the 1980s CIA funding of Bin Laden was a necessary condition for the 9/11 attacks, i.e. that (using our earlier notation) $E(C_0 \& W) = pr(9/11$ attacks without the CIA funding$) = 0$. But quite consistent with that, it may also be, indeed in this case likely is, that even with the CIA funding the probability of the 9/11 attack was still pretty low at that stage. In symbols, $E(C_1 \& W) = 0.1$, say. Thus, applying (1), the CS of the CIA funding is yielded by:

$$E(C_1 \& W) - E(C_0 \& W) = 0.1 - 0 = 0.1$$

That is, despite being necessary, the CIA funding of Bin Laden had only a low CS. Hence necessity is not *sufficient* for high CS.[12]

Turn now to why it is also not *necessary*. Imagine any factor A and effect E such that, in obvious notation, $E(A_1 \& W) = pr(E \text{ given } A_1) = 0.9$, and

11. Admittedly, both my counterexamples will require indeterminism. But cases that are not indeterministic in the relevant sense, i.e. indeterministic relative to our causal knowledge, are likely to be few and far between in the study of history. Besides, for our purposes all we need is that there are *some* indeterministic cases.

$E(A_0 \& W) = pr(E \text{ given } A_0) = 0.1$. Then from (1), the CS of A = 0.9 – 0.1 = 0.8. Thus A has a high CS, yet because $pr(E \text{ given } A_0) = 0.1 > 0$, so A is not necessary. Thus an unnecessary factor may have high explanatory importance.[13]

More sophisticated approaches, such as those of Martin or Hammond, modify the necessity analysis a little. For instance, following (Martin 1989, 78), if A and B are two causes of P relative to Q, where Q is a "comparison situation" (roughly, E_0 in our notation), then:

> A is a *more important* cause of P relative to Q than was B if:
>
> 1) either A was necessary for P or B was not necessary for P
> 2) had B not occurred, something would have occurred which more closely approximates P than had A not occurred.

In our notation, condition 2 might be expressed: $E(B_0 \& W)$ is closer to P than is $E(A_0 \& W)$. This is related to the issue of formally incorporating E_0-sensitivity, mentioned in footnote 8 earlier. I endorse such incorporation (although not necessarily this particular version of it). But I shall focus here on condition 1 instead. That condition still implies that under no circumstances can an unnecessary cause be more important than a necessary one. But our counterexamples have just demonstrated why this is false. An unnecessary cause raising the probability of the effect from 0.1 to 0.9, for instance, may certainly be more important than a necessary cause that merely raises it from 0 to 0.1. Generally, the fact that explanatory importance can come in differing degrees tells against any identification of it with necessity.

4. Conclusion

There is value in being explicit about the notion of explanatory weight. We see that weighted explanations need to be clear on the intended contrast classes to both cause and effect, as well as on background conditions. That is, in this article's notation, they need to be clear on all of: C_1, E_1, C_0, E_0, W.

12. Similar analysis is possible for many 'deep' underlying causes. For example, even if human greed and tribalism were necessary for the Second World War, were they also explanatorily important of it? If the contrast case E_0 is merely some slightly different war, then probably not.

13. This last point directly contradicts, for instance, the view of the nineteenth-century German theorist von Kries. Von Buri and von Bar, the influential legal theorists from that time, were also advocates of a necessity analysis of CS (Turner 1986).

Then we must try as best we can to evaluate those counterfactuals and probabilities highlighted by formulas (1) and (2).

Are such evaluations possible? Being explicit shows us just how hard are the epistemological requirements for weighted explanations, especially in history. Accordingly, assigning causal-explanatory weights is difficult and thus potentially controversial. Indeed, perhaps this very difficulty helps explain historians' frequent recourse to *narrative* explanations instead. The latter tend to adopt descriptions that automatically import much background knowledge relevant to the evaluation of causal strengths.

My own view is that the prospects are not hopeless everywhere, at least not if we are content with approximate evaluations. I would argue that the examples cited in this article, for instance, are sufficiently evaluable to illustrate the salient conceptual points. But readers will form their own views of the matter. This article has aimed rather to provide the necessary precursor to any such consideration of the epistemological situation, namely a prior clarification of the conceptual one.

References

Achinstein, P. 1983. *The nature of explanation*, Oxford: Oxford University Press.

Carr, E. H. 1962. *What is history?* Alfred Knopf: New York.

Lewis, D. 1973. Causation. *Journal of Philosophy* 70:556-67.

Martin, R. 1989. *The past within us*. Princeton, NJ: Princeton University Press.

Northcott, R. 2005a. Comparing apples with oranges. *Analysis* 65 (1): 12-18.

———. 2005b. Pearson's wrong turning: against statistical measures of causal efficacy. *Philosophy of Science* 72 (5): 900-12.

———. 2006. Causal efficacy and the analysis of variance. *Biology and Philosophy* 21 (2): 253-76.

———. (forthcoming). Apportioning explanatory responsibility.

Pearl, J. 2000. *Causality*. New York: Cambridge University Press.

Przeworski, A., and M. Wallerstein. 1988. Structural dependence of the state on capital. *American Political Science Review* 82 (1): 11-29.

Sober, E. 1988. Apportioning causal responsibility. *Journal of Philosophy* 85:303-18.

Sober, E., E. O. Wright, and A. Levine. 1992. Causal asymmetries. In *Reconstructing Marxism*, 129-75. London and New York: Verso.

Spirtes, P., C. Glymour, and R. Scheines. 2000. *Causation, prediction, and search* 2nd ed. Cambridge, MA: MIT Press.

Turner, S. 1986. The Search for a methodology of social science. Dordrecht: Reidel.

———. 2007. Causation in history. In *The Blackwell guide to the philosophy of history*, edited by Avi Tucker. Oxford: Blackwell.

Van Fraassen, B. 1980. *The scientific image*. Oxford: Oxford University Press.

Woodward, J. 2003. *Making things happen: A theory of causal explanation*. New York: Oxford University Press.

Robert Northcott is an assistant professor in the Department of Philosophy at the University of Missouri-St Louis. He works primarily on causation, in particular on the related issues of its appliance to science and the metaphysics of what it is. He has published several articles in this and other areas, including the philosophy of economics. For more details, visit: http://www.umsl.edu/~northcottr/.