CHAPTER 7

PARTIAL EXPLANATIONS IN SOCIAL SCIENCE

ROBERT NORTHCOTT

7.1. Introduction

How much was the increased murder rate explained by higher unemployment? What was the main cause of the American Civil War? Was it the penetrating offense or the stout defense that was most responsible for the football team's victory? It is ubiquitous in social science and indeed everyday life that the causes we have identified explain some but not all of an outcome. In such cases, the question of critical interest is to quantify each cause's contribution to the outcome. The focus is not on how general or deep or transportable a particular explanation or mechanism is, important though those concerns may also be, but rather is narrowly on how much a cause explains an effect in a particular one-off case. This is relevant historically to determine which factors explained an outcome most. It is also relevant as a guide to future intervention—which factors *would* influence an outcome most?

Comparing different causes' importance, and apportioning responsibility between them, requires making good sense of the notion of partial explanation, that is, of *degree* of explanation. This turns out to be a delicate task.[1] The vast literature on defining causation itself is of no direct help because in the cases of interest here typically all parties already agree on what causes are present. The issue at hand is, rather, degree of causation, which is clearly distinct from mere causation *simpliciter*. It turns out to be very useful to make our concepts in this area explicit. What do partial explanations amount to, and, thus, what constitutes good evidence for them? How much is degree of causation subjective, how much objective? If the causes in question are probabilistic, how much is the outcome due to them and how much to simple chance? What is the role of contrasts?

OUP • UNCORRECTED PROOF

This chapter is split into four sections. Throughout, the emphasis will be primarily conceptual rather than epistemological. I begin by formulating the notion of degree of causation, or *effect size*. One particular understanding of this is standard across many sciences, and I relate it to influential recent work in the literature on causation. I use this understanding as the basis for my understanding in turn of *partial explanation*. In the second section, I examine to what extent mainstream social science methods—both quantitative and qualitative—succeed in establishing effect sizes so understood. The answer turns out to be, roughly: only to some extent. Next, the standard understanding of effect size, even though widespread, still has several underappreciated consequences. In the third section, I detail some of those. Finally, in the fourth section, I discuss the separate issue of explanandum-dependence, which is essential to assessing any cause's explanatory importance and yet which has been comparatively neglected.

7.2. Degree of causation

Let X be a cause variable and Y an effect variable. Y is a function of the state of the world, that is of X and W, where W is background conditions (i.e., formally a set of variables representing the state of the world just excluding X).[2] Let x_A denote the actual value of X, and x_C the salient counterfactual value of X. And let y_A and y_C denote the values that Y takes given x_A and x_C respectively (given actual background conditions).[3] Then define a *causal effect* (or, equivalently, the *effect size* or *strength* or *importance*) of a cause variable X with respect to an effect variable Y, to be:

$$y_A - y_C \qquad [1]$$

Formula [1] is quite intuitive—in words, it is how much difference X made. Glossing over various technical details, we are interested in the quantity of effect for which x_A is responsible, and this is just the level of effect with x_A compared to the level with some alternative input x_C. 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.[4] A negative CS here would correspond to accelerating the ball backward; a zero CS to leaving its acceleration unchanged. The units of a CS are whatever units are used to measure Y. At the heart 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 then yield only the combined impact of the two changes. For this reason, in [1] the background conditions must be constant across the two terms.[5]

OUP • UNCORRECTED PROOF

y_C, the right-hand term in [1], is a counterfactual—we are interested in what the level of effect *would* have been, given x_C and background conditions. How can this term be evaluated? Because, in reality, background conditions are never quite exactly the same from moment to moment, epistemologically the best we can ever do is find data from as good a re-creation of them as possible. In this respect, [1] serves as a normative ideal, guiding our treatment of actual data by telling us what hypothetical quantity is relevant to evaluating a CS. Only some actual data, namely those adequately approximating controlled constant-W conditions, will then be appropriate.[6]

Once we know a given cause's CS, we may compare it to the strengths of other causes. In this way, in principle we can compare several causes to see which of them are more responsible for a particular outcome. As is widely recognized, this notion of relative importance is choice-of-sample-specific. For example, a gas leak may be deemed a more important cause of a fire than is the presence of oxygen because of an implicit relativization to a normal population of cases. More particularly, relative to that normal population the unusual, difference-making feature is the gas leak not the oxygen. Formula [1] (and, later, formula [3]) captures this sensitivity via its explicit relativization to choice of contrasts. Change the salient contrast situation and you change which cause is fingered as having a large CS.[7] (This reflects the highly context-sensitive nature of any CS—see section 3-1.)

Not surprisingly, [1] or something like it has a long history in several different literatures as a measure of degree of causation. In the philosophy of history, 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; Northcott 2008a). More recently, measures in psychology, psychiatry, epidemiology, law, and computer science are also similar. Moreover, still other measures are closely related, being again essentially comparative of an effect with and without a cause.

There are many possible formulations for [1] other than a simple difference. Why not a ratio, for instance? But [1] is more flexible than might first appear; in particular, it may be calculated anew for any choice of Y-variable. So a concern with ratios, for instance, is readily accommodated by applying [1] to the logarithms of the original Y-variable. In this way, odds ratios too, for instance, are also causal strengths in the sense of [1]. Moreover, often we are interested in the *variance* rather than level of a variable (Braumoeller 2006); after suitable re-definition of Y, again [1] applies straightforwardly.

Within analytic philosophy, [1] reflects the common emphasis 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. The form of [1] can be incorporated into the contemporary Bayes net and causal modeling literatures, and arguably is endorsed by experimental practice, at least in the case of quantitative variables (Woodward 2003; Pearl 2000; Spirtes, Glymour, and Scheines 2000). More generally, it is also consistent with the mainstream literature on probabilistic causation (Hitchcock 1996). Finally, it also represents the impact of a hypothetical *intervention*, in particular one changing X from x_A to x_C, reflecting a close affinity with a manipulationist view of causation (Woodward 2003).

OUP • UNCORRECTED PROOF

Both terms in [1] can be interpreted as expected values. And the Y-variable may be a probability.[8] In these ways, [1] accommodates the possibility of indeterminism. The indeterminism here is uncertainty regarding what effect results from a particular specification of cause and background conditions.[9] In practice, such indeterminism is ubiquitous in social science; its accommodation is therefore essential for any definition of CS.

There are few alternative accounts of CS with any currency. In practice, much the most widespread in science are those derived from the analysis of variance and a range of related statistical techniques. But the many critiques of these techniques as instruments for assessing causal responsibility are by now familiar. For a sampling, see Lewontin 1974; Shipley 2000; Spirtes, Glymour, and Scheines 2000; Northcott 2008b. Other alternatives will be discussed below, in particular regression coefficients and various nonprobabilistic qualitative definitions, as well as the apparent similarity between [1] and measures of average causal effect in statistics. Also, the final section below will discuss an issue—explanandum-dependence—that does suggest [1] needs to be augmented.

7.3. Causal Strength and Social Scientific Practice

7.3.1. Two traditions

Two different methodological traditions have arisen within political science, which Mahoney and Goertz (2006) label "two cultures." Analogous traditions exist in other social sciences too. These traditions are:

1. *Qualitative*. This typically examines the causal structure behind only a small number of events, via detailed historical investigation. The structure is taken to apply to every event in the sample, thus allowing qualitative comparative analysis ('QCA'). (Ragin 1987 is an especially influential advocate.) For example, what explains social revolution in agrarian-bureaucratic states that were not formally colonized (Skocpol 1979)? At the time of Skocpol's work, this analysis applied to only three examples, namely the revolutions in France, Russia, and China. In Mahoney and Goertz's phrase, the focus is on "causes of effects"—what is the detailed causal structure behind these particular events?
2. *Quantitative*. This typically analyzes conditioned statistical associations over large samples, often by running regression analyses. For example, what is the influence of earnings and education on political participation? The focus is now on "effects of causes"—what is the average level of effect when a cause is present compared to when it is absent? King, Keohane and Verba (1994) are especially influential advocates of the quantitative approach.

OUP • UNCORRECTED PROOF

I will focus on the specific issue of how each tradition furnishes us with partial explanations. That is, do their methods successfully yield us causal strengths?

7.3.2. The qualitative approach

A theory in the QCA tradition may be summed up by a simple Boolean formula. Take, for instance, Moore's (1966) theory of the development of early modern democracies. Expressed in Boolean terms:

$$Y = X \,\&\, (A \vee B)$$

where Y = democracy, X = a strong bourgeoisie, A = alliance between bourgeoisie and aristocracy, B = weak aristocracy. In words, what was necessary for the development of democracy was a strong bourgeoisie, together with an aristocracy neutered either by its own weakness or by alliance with the bourgeoisie.

There are several characteristic features to note:

1. The theory offers a causal *structure*, not just a singular causal explanation, typically with *multiple pathways* to the same outcome.
2. It is meant to apply to *several* cases.
3. It is *deterministic*, couched in terms of necessary and sufficient conditions.

These features lead in turn to two noteworthy consequences: First, causal explanations are provided of singular instances. The Boolean formula, assuming it is correct, provides in particular instances an appropriate *invariance relation* (Woodward 2003)—that is, it licenses answers to the what-if-things-had-been-different questions characteristic of causal explanations. For example, given that there was a weak aristocracy, then the bourgeoisie being strong rather than weak explains why there was democracy rather than no democracy.

Second, however, it only enables us to calculate causal strengths, and thus partial explanations, that are trivially all-or-nothing. For example, a strong bourgeoisie receives 100 percent credit because it is a necessary condition. According to formula [1], nontrivial causal strengths are possible in such cases only if the analysis is probabilistic. But the qualitative approach, at least in its QCA version, is characteristically and explicitly nonprobabilistic.[10] This unfortunately renders it inapplicable to the many explananda for which we do not have full explanations—that is, for which the causes we identify account for only some rather than all of the effect.

Aware of this limitation, several ingenious recent papers attempt to evade the second conclusion, and to define nontrivial causal strengths even within a framework of deterministic necessary and sufficient conditions. I applaud this recognition of the problem. Although there is no space here to discuss them fully, I will very briefly outline some of these recent proposals before indicating why I think they are unsatisfactory substitutes for [1].

Mahoney, Kimball and Koivu (2009) present a definition based partly on set theory and partly on sequence analysis. If X is necessary for Y then on a Venn

OUP • UNCORRECTED PROOF

diagram Y is a subset of X, and vice versa if X is sufficient. Matters of degree may then be incorporated via an implicit metric over the Venn diagram space. For instance, if X is a sufficient condition for Y, it may be more or less important depending on how much of set Y is filled by set X, that is, depending on the size of (Y – X) relative to X. Sequence analysis, roughly speaking, then traces the impact on such set-theoretic calculations of adding in new entries to the chain of causes leading up to the effect. In this way, a qualitative importance ranking can be defined over the different causes in the chain.

Despite its ingenuity, the Mahoney, Kimball, and Koivu proposal has some weaknesses. Perhaps the most striking are two of scope. First, the authors themselves note that their scheme is able only to compare different causes on the same chain. It is therefore inapplicable to causes that are not themselves causally linked, that is, not to multiple independent or simultaneous causes. Second, there is a further serious restriction of scope, namely inapplicability to quantitative explananda. By the latter, I do not mean here the large-sample regression analyses mentioned already. Rather, I have in mind singular cases where the effect of interest is not a simple all-or-nothing event, such as a revolution, but rather a quantitative variable. For instance, what explains the rise in average global temperature over the last few decades? This effect is a sample of one, but we will typically be interested in the quantitative contribution to it of various factors—how much did deforestation contribute, how much did air travel, how much did volcanoes? Similar remarks apply to many other effects of interest, such as crime rates, voting shares, or economic growth indices.

In addition to these restrictions of scope, there are significant conceptual worries. Mahoney, Kimball, and Koivu's scheme requires a metric over Venn-diagram space, but is not clear either how to motivate such a metric or about the source of the particular universe of possibilities over which the sets range. Yet these issues are crucial to their scheme. Ultimately, they relate to choice of comparison population or, in the terms of [1], to choice of contrasts, and should be addressed explicitly. Finally, whereas [1] is easily interpreted as representing the result of an intervention, results from Mahoney, Kimball, and K. Koivu's scheme—because framed set-theoretically—can only be interpreted as degrees of necessity or sufficiency. It thus cannot inform us precisely about the impact of a future or past hypothetical intervention. (This shortfall is related to the inapplicability to quantitative explananda.)

A second ingenious nonprobabilistic definition of CS for singular cases is due to Braham and Van Hees (2009). Much of their focus is on the relation between causal and moral/legal responsibility, but their measure is intended to apply to the former. They import tools from game theory and voting theory to define a cause's strength to be, roughly speaking, the proportion of possible routes to the effect that include that particular cause. Unlike Mahoney et al's, this new measure certainly can be applied to multiple independent causes. On the other hand, this gain is offset by it now no longer being clear that it can be applied to different causes on a single chain. ([1] can be applied in both circumstances.) More fundamentally, Braham and Van Hees's approach suffers from the same shortfall as Mahoney et al's, namely in

OUP • UNCORRECTED PROOF

being inapplicable to quantitative effects.[11] It also is subject to an analogous worry as before regarding interpretation of their results.

Ragin (2006) and Goertz (2006) present a third approach, framed in terms of fuzzy sets—that is, sets whose membership can be a matter of degree. For instance, a state might be deemed a democracy to degree 0.6. Naturally, this makes it much easier to define a nontrivial measure of CS even for the qualitative case, and moreover a measure that is applicable to multiple independent causes. On the other hand, the particular measures endorsed by Ragin and Goertz seek to define one summary CS score for a whole population (see below for why this is problematic).

Goertz (2006) also analyzes CS in terms of actual frequencies. In particular, when considering the CS for a necessary condition in a particular one-off case, he appeals to a wider sample of analogous cases. For instance, suppose state breakdown is necessary for social revolution but only actually leads to it on 3 of 16 relevant occasions (Goertz 2006, 92). Then, roughly speaking, it is assigned a CS of 3/16. Formula [1] appeals to probabilities; if those probabilities are interpreted as actual frequencies rather than single-case chances, then (assuming agreement on the relevant sample) it would yield the same result as Goertz's scheme (bracketing for now issues about background conditions). On the other hand, three worries: First, the actual-frequency approach leaves us vulnerable to actual samples that do not happen to proxy satisfactorily the counterfactuals relevant to CS. Second, Goertz's scheme again seeks to define a single score for CS across a whole population of cases. And third, it seeks to combine into a single measure not only CS as understood in this chapter but also what Goertz calls "trivialness"—by the latter Goertz has in mind something like gravity, which may be a necessary condition for revolution but, since it occurs in all relevant cases of nonrevolution too, is of only trivial interest and thus should have its CS downgraded. But, in line with philosophical consensus, my view is that such trivialness should lead merely to our pragmatic neglect of gravity rather than to a denial of its causal strength (see the references in note 7 for discussion).

Finally, none of these proposals speaks to the range of issues that [1] does. In particular, they address either incompletely or not at all the matters to be discussed in sections 3 and 4 below.

7.3.3. The quantitative approach

I will focus on the relation between causal strength and regression coefficients. To keep the conceptual issues clear I will consider just the simplest case, that is, a regression equation of the form:

$$Y = a + bX + e$$

where Y is the dependent variable, X the independent variable, a and b are constants, and e is an error term. For instance, perhaps Y = political participation, and X = earnings. The relevant population will typically be large, for example, every adult citizen.

OUP • UNCORRECTED PROOF

The regression coefficient b is usually taken to represent the influence of X on Y. The central question is: Does it represent X's causal strength?[12] Start with the simple binary case where X only takes two values, as when a population is split into treatment and control groups. The coefficient b is then a measure of X's causal effect, namely:

$$Y(X = 1) - Y(X = 0) \qquad [2]$$

where 1 and 0 are conventional values for treatment X being present or absent, and where the values for Y are averages across all the data points in the sample. Thus the quantitative tradition's conception of causal effect, [2], is obviously analogous to the CS of [1], albeit now applying to particular populations of individuals rather than to those individuals themselves (on which more shortly).[13]

However, there is also one obvious and significant disanalogy between formulas [1] and [2], namely the issue of *confounders*. In particular, the right-hand term Y(X = 0) in the quantitative formula is derived from actual data. The hope is that these data serve as satisfactory proxies for the counterfactuals deemed relevant by [1], but there is no guarantee of that. In particular, they will not do so if there is too much causal heterogeneity within the population. For instance, suppose lack of education causes people both to be low-earners and also not to participate in politics. Then there will be a correlation between the latter two variables, reflected in a positive score in [2], even if there is no causal relation between them because the real work is all being done by lack of education. Formally, the background conditions are inconstant in such a way that the actual data from the X = 0 ("control") group are not a good proxy for the counterfactual of what would have happened to the X = 1 ("treatment") group *if* the latter's members had not received the treatment.

The problem of confounders is avoided by [1] because the right-hand term is a counterfactual relativized to constant W, precisely so as to avoid spurious correlations and thus spurious CS scores. The problem is also addressed in the paradigm case of a randomized controlled trial, where all confounders are (at least on average) equally distributed between the treatment and control groups. But most regressions are run on observational rather than experimental data. Accordingly, in those cases there is no guarantee that [2] captures the size of a genuinely *causal* effect.

Of course, social scientists are long familiar with this issue and a huge literature has built up around how best to test for and avoid such confounding. (Morgan and Winship 2007 and Pearl 2000 are notable recent contributions.) Much boils down to sufficient knowledge of the underlying causal structure that is generating the data, as this in turn enables us to ensure there are no significant omitted variables, that the causal relations between the modeled variables are specified correctly, and so on. (Ultimately, there's an analogous requirement in the qualitative case too.) In practice, much of the hard work in social science lies precisely in establishing this requisite knowledge. However, I will pass over the many details of that work here, in order to focus instead on a separate question: If good circumstances do prevail—that is, if the regression equation is indeed a correct causal specification—are

OUP • UNCORRECTED PROOF

regression coefficients good proxies for CS even then? For it turns out that even in such good cases, difficulties still remain.

First, a regression can only estimate a CS whose inputs the regression's particular sample happens to proxy. For example, what if we wanted to know the CS of the actual treatment compared to a dosage outside the actual range? No data from the actual sample would proxy the relevant counterfactual; neither would the actual data alone license any extrapolation of the regression result to a new range of data.[14] (Remember, most analyses are of observational, not experimental, data.)

Second, return temporarily to the case where X is binary, that is, where it only takes two values in the population. For $Y = a + bX + e$, the coefficient b may then be interpreted in either of two different ways:

1. As a fixed constant, giving the CS that applies to the two values of X for every member of the population.
2. As some average of the varied CS scores that apply to each individual member of the population.

These two options are often not distinguished but their implications are significantly different. In particular, in the case of 2, the value of b does not directly bear on individual cases at all. Rather, it gives a CS only at the population level.

(In passing, a note on levels: In accordance with difference-making views of causation generally, formula [1] is not reductionist. That is, [1] may apply to X and Y variables at any level. In particular, X and Y may be population-level variables in which case a positive CS indicates a population-level cause, specific to the particular population sampled (and salient counterfactual population). Where there can be a problem is if background theory yields the causal structure only at the individual level, because then aggregation issues mean it may not be straightforward to infer the relevant causal structure at the population level—and without the latter, it is not possible to identify population-level causal strengths reliably. For instance, it is one thing to have in mind a mechanism for why an individual's education might increase his or her level of political participation. But at the population level, what matters is how *rates* of education affect *rates* of political participation. Perhaps, for instance, after a while ever-higher rates of education no longer much increase rates of political participation because of some macro-threshold effect. The point is that such a population-level mechanism could be invisible to the postulated individual-level mechanism, meaning that we did not have the population-level causal structure right, meaning in turn that we were unable to evaluate the counterfactuals necessary to any calculation of CS. In this chapter, though, I will again focus on conceptual rather than such epistemological issues.

Third, what if, as in the general case, the X-variable is interval-valued rather than binary? Then X may take any of a range of values, and the use of regression coefficients to estimate causal strengths becomes considerably more complicated. For such a population will contain within it many different causal strengths, corresponding to different pairs of values of X. For instance, the data could be used to

OUP • UNCORRECTED PROOF

estimate the CS of X = 2 rather than 1, or of 5 rather than 2, or of some mixture of 3 and 1 rather than some mixture of 2 and 1 . . . and so on. Any regression coefficient gives just a single value for the population and so could only be a melange of these various possibilities.[15,16]

However, a regression equation can still be useful for calculating causal strengths sometimes. For in good cases it will represent accurately the underlying causal structure, and estimation of its coefficients will then enable particular causal strengths to be inferred, much as they can be inferred in physics if we know the relevant law. This may occasionally be especially straightforward. In particular, in the linear case, when X is a cause of Y, and when background conditions were not causally heterogeneous, the value of a regression coefficient will equal the average value of a "unit CS" in that population.[17]

Finally, there are several parallels with the qualitative case. In both traditions, causal strengths can be calculated only if the correct underlying causal structure has been identified. In the one case, this is done via detailed historical investigation, in the other via a mix of this or more abstract theory (in order to motivate the initial regression equation) combined with statistical analysis. Moreover if, against QCA practice, we interpreted qualitative analyses probabilistically, the resultant uncertainty would be analogous to the error term in a regression equation.[18]

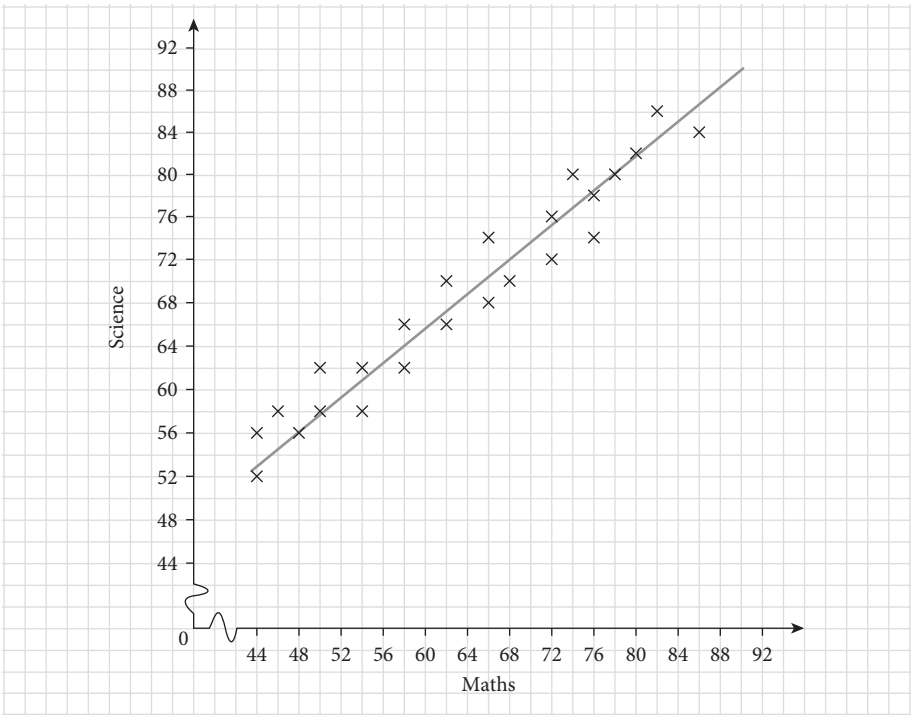

Figure 7.1 A line of best fit Comparing any two data points generates a new CS each time. Many different causal strengths may thus be found within the same population.

OUP • UNCORRECTED PROOF

7.3.4. Recap

Qualitative analysis is certainly capable of delivering CS scores, assuming that the causal structure has been identified correctly. However, QCA's deterministic emphasis purely on necessary and sufficient conditions means that these CS scores will only ever be trivially all-or-nothing.

In favorable circumstances, quantitative analysis too can estimate the correct causal structure, from which, in principle at least, nontrivial CS scores of interest can then be inferred. On the other hand, even in these favorable circumstances, the value of the regression coefficient itself will only rarely equal that of the CS of interest. Moreover, the CS estimates will only ever be at the population; it is then a substantive further issue whether they apply also at the individual level.

7.4. Features of Causal Strength

7.4.1. Context-specificity

Although, as noted, formula [1] is ubiquitous across many sciences, nevertheless several of its implications are often unappreciated. Begin with a quick theoretical detour. According to [1], any CS value for a cause X is relativized to three things:

1. the choice of effect variable Y
2. background conditions (reflecting the levels of other causes of Y)
3. the choice of contrast x_C

Regarding (1), for instance, kicking a ball may be a strong cause of the ball's acceleration but a weak cause of its changing color. Regarding (2), striking a match is a strong cause of light if background conditions include sufficient oxygen in the atmosphere, but not otherwise. Regarding (3), how much difference a cause makes depends on what we are comparing it to. The difference made by a substitute in a sports team, for instance, depends critically not just on the substitute but also on which player they are replacing.

Combining (2) and (3) above, for any given cause-effect type pair there will therefore be many associated token causal strengths—a new one for every change of background conditions or x_C. In a phrase, any CS is *context-specific*. No cause has a univocal strength *simpliciter*. Rather, as background conditions or choice of x_C vary, the very same cause X may have a different strength even with respect to the same effect Y.

Stated baldly like this, the context-specificity thesis perhaps seems so obvious as to be uninteresting. Yet this feature of CS turns out to have several significant implications. One, for instance, is the familiar problem of extrapolation between populations. Any population-level CS is relativized to a particular population

OUP • UNCORRECTED PROOF

(formally, to a particular x_A, x_C and value of W), so there is no guarantee that the same CS will be true also of some new population. Knowledge of the underlying causal structure will, of course, be relevant to determining whether the CS value can indeed be so extrapolated.

Other implications of context-specificity seem to be less widely recognized. I survey several of these now. (Northcott 2008a covers some of the same ground in more detail.)

7.4.2. Nonadditive Causal Interaction

Consider a fictional example, simplified for illustration. Suppose that, statistically, graduating from college has a certain impact on earnings for whites—say, it increases it by four units (in comparison to finishing education before college). For blacks, on the other hand, college raises earnings by only one unit. In other words, race and education interact nonadditively, as summarized below:

	No college	College
Black	2	3
White	3	7

What is the CS of college education here? *There is no univocal answer*. It is four units for whites, one unit for blacks, and presumably some weighted average of those two numbers for the population as a whole (assuming both that there are no other races in the population, and that the figures in the table are good proxies for the counterfactuals relevant to CS calculations). Moreover, the same is true for the CS of race, which may be either one unit or four. That is, being white rather than black increases earnings by a different amount, depending on whether one has a college education. This lack of univocality follows directly from the context-specificity of any CS—in general, there is a different CS for every change in background conditions.[19]

Notice an important feature here. Consider the income of a college-educated white. The CS of their college-rather-than-no-college education is four units; that of their being white rather than black is also four units. These two causal strengths therefore add up to *more* than their total actual earnings of seven units. But there is no contradiction here. Neither individual CS is more than the total actual effect. Arguably, the adding up to more than the total is just what we should expect—that is the very meaning of nonadditivity. It is a mistake to think of a fixed pie of causal credit waiting to be divided up between the causes. Rather, it may be that many causes simultaneously have a large CS, or that none does. As it were, the total amount of pie can vary.[20]

How does current social scientific practice handle interactive effects? Begin with qualitative approaches. First, Braham and Van Hees's (2009) scheme to define nontrivial causal strengths in a deterministic setting is committed to a fixed-pie view of causal credit, which is problematic in nonadditive situations, for the reasons just noted. (Mahoney, Kimball, and Koivu's (2009) scheme, recall, is not applicable

OUP • UNCORRECTED PROOF

to multiple independent causes in the first place.) By contrast, and for all its other difficulties, the traditional QCA apparatus of necessary and sufficient conditions does take on board nonadditive interaction appropriately. For instance, two independently necessary conditions—which is formally a case of positive interaction—will both be assigned full causal importance by it.[21]

What of regressions? Often, one of the independent variables in a regression equation is an interaction term, precisely in response to the possibility that two other independent variables interact nonadditively. As noted earlier, sometimes regression estimation can help establish the causal structure from which a CS of interest may then be inferred even though the regression coefficients themselves typically do not correspond to CS values. The same applies now in the nonadditive case. In general, the size of an interactive effect between two causes is the CS of both causes together, minus the causal strengths of each cause individually (hence is of size three units in our example above, as per note 20). There is no reason to suppose the regression coefficient for an interactive term will track this quantity, and besides there will in any case be many different values for such an interactive effect within a population just as there are many different values for a CS. For the purpose of elucidating causal strengths, therefore, inclusion of an interaction term is of dubious relevance.

7.4.3. Small Causes can be Important

At Waterloo, the British fought Napoleon to a standstill; the arrival of the Prussians was merely the straw that broke the camel's back. Is it true that the British contribution was therefore more causally important? Measurably asymmetric contributions, the reasoning runs, imply asymmetric causal importance. But it turns out that this is not true in general—even if we do 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.

To see this, for simplicity assume determinism here. Label: x_A = the fighting of the British soldiers, as measured by their number; z_A = the fighting of the Prussian soldiers, as measured by their number. Let the salient contrasts in both cases be zero soldiers. Denote Napoleon's actual defeat by Y = 1, his nondefeat by Y = 0. Finally, in line with historical consensus, assume that neither the British nor the Prussians could have won the battle alone.

Then, applying formula [1]:

1. CS of the British = the outcome with the British compared to that without them = (Y given x_A and z_A – Y given x_C and z_A) = 1 – 0 = 1
2. CS of the Prussians = the outcome with the Prussians compared to that without them = (Y given z_A and x_A – Y given z_C and x_A) = 1 – 0 = 1

That is, the British and Prussians are awarded equal causal strengths with respect to winning the battle. This is true regardless of the exact number of soldiers each contributed—the calculation would have been unchanged even if the British had ten

OUP • UNCORRECTED PROOF

times as many men as the Prussians. The result follows (in this deterministic case) simply from x_A and z_A's individual insufficiency and joint sufficiency.

With respect to many *other* effects, of course, the British would have had far greater causal importance because of their far superior numbers. Probably too our moral intuitions may be sensitive to such "type" considerations. But with respect to this particular effect in these particular circumstances, we have no justification for awarding the British any more causal credit than the Prussians. Besides, with respect to some other effects, say the army's German-language competence, it is the Prussians that would be more important. Any CS is a token, sensitive to every change in effect variable as well as background conditions.[22]

One source of confusion is the easy *commensurability* of the number of British and Prussian soldiers. This fools us into thinking that the British must therefore somehow have greater CS here. But such commensurability is a red herring: Commensurable or not, the British and Prussian forces were nonetheless equally and symmetrically necessary to the defeat of Napoleon at Waterloo. More generally, commensurability can be shown to be neither necessary nor sufficient for asymmetric attributions of CS (Northcott 2005a).

The take-home message is that even so-called small causes, such as the Prussian army at Waterloo, can have big causal strengths. Indeed, more fundamentally, it seems to me that what it even means for a cause to be large or small can only be a large or small CS. It follows that a cause's size can only be defined relative to a choice of *effect* variable, not by considering the cause variable in isolation from any explanatory context.

7.4.4. Limits versus Selections

In the next two subsections, there is space briefly to flag two other common confusions. (See Northcott 2008a for further discussion.) First, a claim often heard, from Marxists and others, is that the deep structural factors underpinning history are the only truly important ones. For example, Przeworski and Wallerstein (1988) argue that the threat available to capitalists of an investment strike, a threat that in the past has often been exercised, has constrained the politically possible range of government policy throughout postwar Latin America. 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. The background limits are what matter, not the subsequent restricted selections.

But it is a mistake to privilege the deep structural factors as necessarily more important (again understanding causal importance, here as elsewhere, to be CS as defined by [1]). 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 important than the threat structure, which narrowed the field merely from 30 to 25.

OUP • UNCORRECTED PROOF

All turns on the precise explanandum of interest. Suppose the actual policy is a conventional one. Then it may well be that 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 the particular choice of minister—that explains why we have the actual policy rather than a *radical* one. But at the same time, it may well be the particular choice of minister—and not the threat structure—that explains why we have the actual policy rather than an *alternative conventional* one. (Northcott 2008a explicates this rather intuitive line of reasoning more formally. See also Sober, Wright, and Levine 1992, from which the example is adapted.) I discuss the issue—crucial in this example—of explanandum-dependence in section 4 below. The point here is that there is no general, explanandum-independent sense in which limits are more causally important than selections.

7.4.5. Underlying versus Proximate Causes

The second common confusion concerns the relative strengths of underlying and proximate causes. In particular, there is no general reason to think the former more important. Intuitively, merely being an earlier link on a causal chain does not imply also being a more important one (Sober, Wright, and Levine 1992). For example, suppose the first link in a chain raises the probability of an effect from 0.1 to 0.2, whereas a later link raises it from 0.2 to 0.7. The second cause is clearly the more important, as [1] reflects. It is also evident the situation could be reversed just as easily. Position on the chain in itself implies nothing.

There is one clear objective asymmetry though: An earlier link on a chain is a cause of a later link, but not vice versa. Does this asymmetry imply greater CS for the earlier link? The 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, 65). But this is a red herring. Formally, the asymmetry is already reflected by the earlier cause (or rather, its consequences) forming part of the background conditions in the later cause's CS calculation, whereas the later cause is completely absent from the earlier cause's CS calculation. But that asymmetry does nothing to negate the central fact that what actually matters for CS is only how much a cause raises an effect's probability at the time of that cause's own occurrence.

Often, a preference for the greater importance of underlying causes reflects a focus on a particular kind of explanandum. In particular, if we are interested in longer-term or broader-scale effects, then with respect to them indeed an underlying cause will often have a higher CS than a shallow proximate one. Perhaps these kinds of explananda are more often of theoretical significance, but ultimately this is still all interest-relative. For example, if we are interested instead in the precise time at which a war broke out, then the war's short-run proximate triggers may well now have a much higher CS than its deeper underlying causes.

OUP • UNCORRECTED PROOF

7.5. Explanandum-dependence

Despite its importance, evident in both the preceding subsections, providing a formal account of explanandum-dependence is a notable lacuna in the philosophical literature. Among the few even to emphasize it in this context are Sober, Wright, and Levine (1992), Martin (1989) and Miller (1987). Rectifying this lacuna turns out to have significant consequences for our understanding of what a cause's explanatory importance amounts to. I turn to the issue now.

7.5.1. Roots in the literature

As noted, formula [1] or something like it has a long history in several different literatures. Yet, notwithstanding this ubiquity, it cannot yet be a complete account of degree of causal *explanation*. The reason is that, at least according to contemporary theory, causal explanation is contrastive with respect to both cause *and* effect. On this view, explanation takes the general form:

x_A-rather-than-x_C explains y_A-rather-than-y_C

where y_A and y_C are respectively the actual and contrast values of the effect variable Y. The contrastive view dates from Dretske (1972). Notable developments of it include Van Fraassen (1980), Garfinkel (1981), Achinstein (1983), Hitchcock (1996), and—most influential recently—Woodward (2003). A major advantage of it is that its sensitivity to y_C enables us, as we will see, to capture explanandum-dependence.

The problem—for everyone—is that formula [1] is insensitive to y_C. To be sure, "y_C" appears in [1], but there it is by definition the value that Y would have obtained given x_C. The y_C that is relevant here, as we will see, is that prespecified by an explanandum quite independently of any x_C. To avoid confusion, I will denote explananda by y_A*-rather-than-y_C*. Thus, in the asterisk notation, the real issue is [1]'s insensitivity to y_C*. Rather, [1] incorporates choice of contrast only on the cause side (i.e., x_C), which is fine for an analysis of causal strength but not for one of *explanatory* strength ('ES').[23] No one (to my knowledge) has ever adapted the full contrastive machinery to the issue of *degree* of explanation. To be clear: y_C*-dependence is already standard in theories of causal explanation; the problem is its absence from the standard definition of degree of causation/explanation.

Consider a tragic example: the first dropping of an atomic bomb in war. What explains the timing of this event—that is, of the bomb at Hiroshima? Consider two causes: the fine weather that day; and Japan's reluctance to surrender. In accordance with contrastive theory, we must also specify respective salient contrasts: bad weather that day; Japan amenable to surrender. Here's the point: *Both* these causes can be argued greatly to have increased the bombing's probability:

1. If the weather had been bad instead of good, the bombing would have been postponed.[24]

OUP • UNCORRECTED PROOF

2. If Japan had been willing to surrender, Truman would likely have thought the bombing unnecessary.

But the two causes increase the probability of, so to speak, different aspects of the bombing:

1. The weather impacted its precise timing.
2. The Japanese attitude impacted only its rough timing (or, for those optimists who believe that there might otherwise never have been such an event, whether a first atomic bombing occurred at all).

Thus the weather is highly explanatory only of the short-term explanandum—why the first bombing occurred on August 6, 1945 rather than the subsequent few days. Japan's attitude, by contrast, is highly explanatory only of a longer-term explanandum—why the first bombing occurred in 1945 rather than some subsequent year. Unfortunately, the simple formula [1] is unable to capture this crucial distinction. Both the weather and the Japanese attitude made a big difference to whether the bomb was dropped, and for this reason both factors score well on [1]. It is impossible to represent in [1] the crucial distinction between the long-run and short-run explananda, because just specifying the bomb-drop alone (i.e., just y_A* alone) still leaves unclear which aspect of it is explanatorily relevant.[25] We need a successor to [1] that is y_C*-sensitive.[26]

Neither is this need of merely theoretical interest. The Hiroshima example illustrates its pertinence, as do the issues discussed earlier of underlying versus proximate causes and limits versus selections. More widely, the worry is that disputes ostensibly about substance often turn out to be cases merely of people talking past each other; or, more precisely, cases of people addressing different explananda and thus not really being in conflict. Sober, Wright, and Levine (1992, 134) comment in this regard: "A problem that constantly befuddles debates about the importance of different causes . . . is the correct designation of the object of explanation (the explanandum)." The point is that these episodes could be avoided by an explicit specification of the explanandum of interest—that is, in our terminology, by a specification of y_C* as well as of y_A*.

7.5.2. A Definition of Explanatory Strength

Generally, providing a formal analysis of ES proves an intricate business. There are several categories of cases and many nuances, but there is space here only to give a bare-bones account of one kind of case, namely where the effect is a qualitative variable with no quantitative nuance. (See Northcott 2011a for more, including the relation between what previous accounts do exist and the one advanced in this paper.)

To begin, any ES will be a relation between an explanandum and an explanans. The sequence of analysis is then as follows:

1. Define a target explanandum y_A*-rather-than-y_C*, this being specified independently of any cause variable X.

OUP • UNCORRECTED PROOF

2. Consider a particular explanans x_A-rather-than-x_C (i.e., an actual level of cause x_A) and a contrast level of cause x_C.
3. Define what it means for the change from x_C to x_A to fully explain that from y_C* to y_A*, then assess to what extent this actually occurs.

Regarding (3), the phrase "fully explain" deserves explication. So, to be clear: The sense of "fully explanatory" I will have in mind is when a cause makes all (rather than only some of) the difference with respect to an effect. This is the sense that is of critical interest when considering interventions. Moreover, any difference-making view of causal explanation naturally lends support to describing as "fully" explanatory any cause that makes all the difference; what else, on a difference-making view, could full explanation be?[27]

More formally then, in the sense just specified, x_A-rather-than-x_C *fully explains* y_A*-rather-than-y_C* if and only if the following two conditions are both satisfied:

1. $pr(y_A^* / x_A \,\&\, W) = 1$ (i.e. y_A*) occurs when x_A occurs
2. $pr(y_C^* / x_C \,\&\, W) = 1$ (i.e. y_C*) *would* have occurred had x_C occurred[28]

Our true goal, however, is a measure of *partial* explanation. We may now define this as the degree to which the conditions for full explanation are satisfied. Formally, the ES of an explanans x_A-rather-than-x_C with respect to an explanandum y_A*-rather-than-y_C* is:

$$pr(y_A^* / x_A \,\&\, W) + pr(y_C^* / x_C \,\&\, W)$$

The higher the score the better the explanation, and a maximum score of 2 indicates full or perfect explanation. Often, an ES score of 1 is the neutral one indicating no explanation at all.[29]

Like [1] earlier, formula [3] is rather more adaptable than might first appear. In particular, it is readily extended to cases where the contrast of interest is a *range* of values. For instance, "why was the budget as much as $2m?" can be represented by y_C* = the event of a budget under $2m. Other forms of explananda can be accommodated similarly, such as a concern only with ordinal outcomes (y_C* = the budget was less than the y_A* value). Analogous remarks apply also to y_A*. That is, the same actual event may be described in many ways, possibly impacting the ES score. In this sense, ES is description-dependent. Fundamentally, formula [3] is only defined once given a prior choice of explanandum, that is, of y_A* and y_C*. One result of this is the flexibility to encompass many different explanatory concerns about the same actual event.

The relation between ES and the more commonly cited CS is illuminated by considering interventions. The CS definition [1], recall, directly tracks the impact of these. In particular, it tracks the impact on Y of a change from x_C to x_A. Things are not quite the same with ES. Rather, we can think of ES instead in terms of the desired *result* of an intervention. In particular, ES tracks how well a change from x_C to x_A will yield the desired change from y_C* to y_A*. That is, CS tracks the impact of an intervention; ES tracks to what extent this impact is the one we want.

We may illustrate the ES formula [3] by applying it to the Hiroshima example. Let:

OUP • UNCORRECTED PROOF

y_A* = the actual dropping of the Hiroshima bomb
x_A = (the occurrence of) fine weather that day, x_C = bad weather that day
z_A = Japanese reluctance to surrender, z_C = Japanese willingness to surrender

Recall the short-term explanandum—that is, why the bomb was dropped exactly when it was. This is represented by y_C* = the bomb was dropped in the few days after August 6, 1945. Intuitively, recall, the fine weather is highly explanatory of this whereas the Japanese reluctance to surrender is not. For brevity, simplify here by assuming pr(y_A* /x_A & W) = 1. Then, applying [3]:

1. Good weather's ES = pr(y_A* /x_A & W) + pr(y_C* /x_C & W) = 1 + pr(given bad weather on 6th August 1945, the bomb would have been dropped in the few days after) = (1 + quite high) = (1 + 0.9, say) = <u>1.9</u>.
2. Japanese attitude's ES = pr(y_A* /x_A & W) + pr(y_C* /z_C & W) = 1 + pr(given Japanese willingness to surrender, the bomb would have been dropped in the few days after) = (1 + quite low) = (1 + 0.1, say) = <u>1.1</u>.[30]

Thus, as desired, the weather but not the Japanese attitude is endorsed as highly explanatory.

As is easily shown, the results are reversed for the long-term explanandum: for a new y_C* = a bomb was first dropped only in subsequent years. Then it is the Japanese attitude, but not the weather, that comes out highly explanatory. Of course, what matters here is not the exact figures but rather only the general point they illustrate—namely that [3] successfully tracks those factors that determine ES.

Overall, although of course there may be ways of formalizing explanandum-dependence other than by [3], nothing (to my knowledge) so explicit or quantitative has yet appeared in the literature. Meanwhile, being so explicit carries significant benefits: It helps prevent pseudo-disputes in which the participants are unwittingly talking past one another. It makes clear the connection to the general philosophical literature on causation and explanation. It makes clear the precise roles here of chance,[31] counterfactuals and contrasts. And it also makes clear which aspects of ES are subjective or interest-relative (i.e., choice of contrasts, and initial choice of effect variable), and which are objective (i.e., background conditions, and the value of ES once given Y, y_C* and x_C). It reveals a range of often overlooked features, namely those discussed with reference to CS in section 3. Further benefits, discussed in Northcott (2011a), include: clarity regarding the relation between qualitative and quantitative cases; a general account of explanatory relevance; and, depending on one's wider metaphysical commitments, also some light shed on the relation between causation and causal explanation.

7.6. Conclusion

Formula [1], I have argued, represents our philosophically best-supported measure of degree of causation, one accepted and adopted, implicitly or explicitly, across many sciences. Yet methods widespread in social science conform with it only very

OUP • UNCORRECTED PROOF

imperfectly. In particular, except in unusually favorable circumstances, regression coefficients do *not* track meaningful CS scores, *even when* the regression equation represents the causal structure truly. Current qualitative methods are also ill suited to estimating CS scores in all but deterministic cases. Moreover, if our focus is on degree of explanation rather than degree of causal effect, then we must also take into account the exact specification of the explanandum, yet no current formal technique does so.

On a more positive note, we can do our best to establish the information that [1] tells us is necessary for estimating any CS score. In particular, we need a good estimate of the relevant counterfactual term, which in turn requires a specification of the relevant cause-contrast and background conditions. Sometimes, as noted in section 2, estimating regression equations may be indirectly useful for this purpose. Meanwhile, if we are interested further in degree of explanation then, as formulated by [3] for instance, we need in addition to be explicit about the target explanandum too.

NOTES

1. Notable pioneers in the field include: Good (1961), Holland (1986), Miller (1987), Sober (1988), Martin (1989), Sober, Wright, and Levine (1992), Strevens (2000), Pearl (2000), Spirtes, Glymour, and Scheines (2000), and Hitchcock and Woodward (2003). See also Northcott (2005a, 2005b, 2006, 2008a, 2008b).

2. In causal graph terms, there are arrows into Y from both X and W.

3. For ease of exposition, as well as using y_A, x_C etc., to denote particular values of a variable, throughout I will also use them to denote particular events that instantiate those values.

4. Often, as with temperature or price level, 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 alternative. Thus in its right-hand term, [1] cites the general formulation x_C rather than absence of X or some such.

5. Strictly speaking, in fact the background conditions *do* vary across the formula because as well as impacting Y, in general a change from x_C to x_A will also impact W too. But for our purposes we may ignore that wrinkle, so long as any change in W is only a consequence of the change in X. The point is that conditioning on W eliminates spurious correlations.

6. If the relevant counterfactuals are vague or indeterminate, then so too will be the corresponding CS. Generally, I do not endorse any particular semantics for counterfactuals here, as the salient locus of philosophical dispute lies elsewhere. See Reiss, Chapter 8, for further discussion of counterfactuals.

7. Much recent philosophical and psychological work has focused on the related issues of why we tend to pick out one or other counterfactual dependency among several as marking "the" cause (Hitchcock and Knobe 2009), or more broadly on why some dependencies tend to be judged "causal" more easily than are others (Woodward 2006).

8. I take the probabilities here to be objective single-case chances (and the relevant expected values to be derived from the distributions of such chances). See note 28 below for further discussion.

OUP • UNCORRECTED PROOF

9. I express no opinion on the further metaphysical issue of whether that uncertainty in turn results merely from the coarse-grainedness of such specifications, or in addition from the world itself ultimately being indeterministic all the way down.

10. QCA is not the only possible qualitative approach, however. The alternatives need not be committed to the same deterministic framing, and accordingly may furnish nontrivial causal strengths. I have in mind detailed narrative or historical investigations of single cases, including process tracing (see David Waldner's chapter in this volume). See also the mention of fuzzy sets below.

11. Tellingly, the only quantitative examples considered in their paper feature critical thresholds, so that an effect only ever occurs all-or-nothing, that is, without any possibility of quantitative nuance.

12. Because my focus is on conceptual rather than inferential issues, I pass over here consideration of the error term or of how well b might have been estimated. (I also do not consider here the important question of how X and Y might best be defined and measured in the first place.) Instead, the issue will be: assuming b has been estimated correctly, would it then give the value of X's causal strength?

13. I am glossing over various technical estimation issues here. Several different formulations of [2] are possible, although all share the same basic difference-making idea (Morgan and Winship 2007, 42–50).

14. An analogous problem afflicts the use of analysis of variance here (Lewontin 1974, Northcott 2008b).

15. In principle, stratification techniques could yield a separate regression coefficient for each pair of X values. But in practice, this is rarely done—or even feasible if, for instance, X is continuous. Moreover, in general there will be many different data points for any given X value, rendering even the fine-grained CS an average over many cases.

16. A separate issue concerns the appropriate counterfactual term in each CS calculation. For instance, if $x_C = 1$, should we assess the value of Y using the actual data point for X = 1, or the value of Y calculated from the underlying function, i.e., now ignoring the error term component to the actual value of Y? I do not address that choice here, but the latter option would mean that a regression coefficient might after all capture a univocal CS, as per the next paragraph in the text.

17. Although even this rosy scenario requires in addition that the X-variable is such that one unit of it makes sense. For example, if X represents discrete political parties or policy options, then a one-unit change in X will have no meaning.

18. One disanalogy between the two traditions is that the distinction between individual- and population-level causal strengths is not important in the qualitative case. As noted, the goal in the latter is to outline a causal structure that applies only to (every member of) a particular prespecified small-N population.

19. Notice, therefore, that the presence of nonadditive interaction in no way renders causal strengths somehow incalculable or ill defined. Rather, it simply emphasizes that the same cause-effect pair can have a different CS for every change in W.

20. Neither is it feasible to define each individual cause's credit minimally, allocating the rest of the credit to a joint interactive effect. (In the example in the text, for instance, that would imply giving race and education one unit of credit each for the college-educated white's income, and a joint interactive effect the credit for the remaining three units [above the baseline outcome of two].) There is no space here for a full discussion of why not but, briefly, it turns out that the amount allocated to interactive effects would then depend unacceptably on the arbitrary choice of which causes are foregrounded for our attention

OUP • UNCORRECTED PROOF

(Northcott 2011b). Further, in effect the approach commits itself to the untenable position that each individual cause has a unique context-independent CS.

21. Note though a further implication of [1], which is that merely being necessary is *not* sufficient for full CS, contrary not just to QCA but also to several suggestions in the philosophy of history literature (Northcott 2008a). For instance, a cause might raise an effect's probability from 0 to 0.1, thereby making it necessary, but nevertheless of much less importance than a second cause that although unnecessary raises the effect's probability from 0.1 to 0.9.

22. Braham and van Hees (2009) endorse asymmetric causal strengths in some voting games analogous to the Waterloo example. But I see their reasoning as involving unwarranted appeal to various type considerations.

23. Explanatory strength, or near-synonyms such as explanatory power, has also been used to refer to many other things, such as an explanation's range, robustness, or degree of integration with wider theory. Ylikoski and Kuorikoski (2010) provide a useful taxonomy, although they discuss CS only very briefly. They do also mention the issue of explanandum-dependence but give no analysis of it.

24. In fact, bad weather at Hiroshima would likely have led to the mission being diverted to a target where the weather was better. So we must interpret bad weather here to cover, say, every other sizeable Japanese city too.

25. Neither, it turns out, can this problem be overcome simply by judicious choice of Y-variable (Northcott 2011a).

26. In general, it also proves problematic that [1] is defined in terms of y_A rather than y_A* (Northcott 2011a).

27. Therefore I am not concerned here with explanations that are partial in the sense of specifying only one cause out of the many that determine any given event. On this latter view, full explanation would apply only when we had an accurate description of *all* an event's causes. But this seems pointlessly to insist on an unattainable perfection and in practice would render no explanation anything other than partial. For instance, that a match was struck seems fully explanatory of that match being alight, without also insisting on the details of the friction between match-head and surface, or of why I chose to strike the match, or of every necessary background condition such as oxygen in the atmosphere, the match not being wet, and so forth.

28. "pr()" here denotes probability. As noted in note 8, these probabilities are to be conceived as objective single-case chances. Such chances are, of course, philosophically controversial, and taking them to be the effect variables deviates from usual practice in the causal modeling literature. In defense: In practice, they are invoked ubiquitously in many sciences but it does not seem common that actual disputes turn on disputes about particular evaluations of them. Besides, of course, other accounts of the metaphysics of probability have their own difficulties too. At any rate, such chances seem to be a presupposition of many claims of CS and ES. If their value really is crucially unclear in any particular case, then so will be the associated CS or ES.

I do not endorse here any particular account of objective chance. The goal is only to explicate ES once *given* the probabilities in an explanandum and explanans, not to explicate those probabilities' underlying metaphysics. See Abrams (chapter 9 in this volume) for an account of objective chance applied to the social sciences.

29. An example may help intuition: Suppose we want to explain y_A* = the water in a cup is liquid, rather than y_C* = it is not liquid. Consider an obviously irrelevant explanans, say x_A = I won a game of tennis this morning, rather than x_C = I lost it. Then ES = pr(y_A* /x_A & W) + pr(y_C* /x_C & W) = 1 + 0 = 1. In other words, 1 is exactly that ES score achieved by the irrelevant explanans, and thus is the "neutral" score. (See Northcott 2011a for fuller discussion.)

OUP • UNCORRECTED PROOF

30. Strictly, the composition of W is not constant here across each term, as sometimes it will incorporate the value of X and sometimes that of Z instead. But the main point goes through regardless.

31. For more on [3]'s treatment of indeterminism, see Northcott 2011a.

REFERENCES

Achinstein, P. 1983. *The Nature of Explanation*. Oxford: Oxford University Press.

Braham, M., and M. Van Hees 2009. "Degrees of Causation." *Erkenntnis* 71: 323–44.

Braumoeller, B. 2006. "Explaining Variance; Or, Stuck in a Moment We Can't Get Out Of." *Political Analysis* 14: 268–90.

Dretske, F. 1972. "Contrastive Statements." *Philosophical Review* 81(4): 411–37.

Garfinkel, A. 1981. *Forms of Explanation*. New Haven: Yale University Press.

Goertz, G. 2006. "Assessing the Trivialness, Relevance, and Relative Importance of Necessary or Sufficient Conditions in Social Science." *Studies in Comparative International Development* 41: 88–109.

Good, I.J. 1961. "A Causal Calculus Parts I and II." *British Journal for the Philosophy of Science* 11: 305–18 and 12: 43–51.

Hitchcock, C. 1996. "The Role of Contrast in Causal and Explanatory Claims." *Synthese* 107: 395–419.

Hitchcock, C., and J. Woodward. 2003. "Explanatory Generalizations, Part II: Plumbing Explanatory Depth." *Nous* 37(2): 181–99.

Hitchcock, C., and J. Knobe. 2009. "Cause and Norm." *Journal of Philosophy* 106: 587–612.

Holland, P. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81(396): 945–60.

King, G., R. Keohane, and S. Verba. 1994. *Designing Social Inquiry*. Princeton, NJ: Princeton University Press.

Lewontin, R. 1974. "Analysis of Variance and Analysis of Causes." *American Journal of Human Genetics* 26: 400–11.

Mahoney, J., and G. Goertz. 2006. "A Tale of Two Cultures: Contrasting Quantitative and Qualitative Research." *Political Analysis* 14(3): 227–49.

Mahoney, J., E. Kimball, and K. Koivu. 2009. "The Logic of Historical Explanation in the Social Sciences." *Comparative Political Studies* 42: 114–46.

Martin, R. 1989. *The Past Within Us*, Princeton, NJ: Princeton University Press.

Miller, R. 1987. *Fact and Method*. Princeton, NJ: Princeton University Press.

Moore, B. 1966. *Social Origins of Dictatorship and Democracy*. Boston: Beacon Books.

Morgan, S., and C. Winship. 2007. *Counterfactuals and Causal Inference*. New York: Cambridge University Press.

Northcott, R. 2005a. "Comparing Apples with Oranges." *Analysis* 65(1): 12–18.

Northcott, R. 2005b. "Pearson's Wrong Turning: Against Statistical Measures of Causal Efficacy." *Philosophy of Science* 72(5): 900–12.

Northcott, R. 2006. "Causal Efficacy and the Analysis of Variance." *Biology and Philosophy* 21(2): 253–76.

Northcott, R. 2008a. "Weighted Explanations in History." *Philosophy of the Social Sciences* 3(1): 76–96.

OUP • UNCORRECTED PROOF

Northcott, R. 2008b. "Can ANOVA Measure Causal Strength?" *Quarterly Review of Biology* 83: 47–55.
Northcott, R. 2011a. "Partial Explanations." Unpublished manuscript.
Northcott, R. 2011b. "Causal Interaction and Symmetric Overdetermination." Unpublished manuscript.
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.
Ragin, C. 1987. *The Comparative Method*. Berkeley: University of California Press.
Ragin, C. 2006. "Set Relations in Social Research: Evaluating Their Consistency and Coverage." *Political Analysis* 14: 291–310.
Shipley, B. 2000. *Cause and Correlation in Biology*. Cambridge: Cambridge University Press.
Skocpol, T. 1979. *States and Social Revolutions*. Cambridge: Cambridge University Press.
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: New York: Verso.
Spirtes, P., C. Glymour, and R. Scheines. 2000. *Causation, Prediction, and Search*, 2nd ed. Cambridge, MA: The MIT Press.
Strevens, M. 2000. "Do Large Probabilities Explain Better?" *Philosophy of Science* 67: 366–90.
Turner, S. 1986. *The Search for a Methodology of Social Science*. Dordrecht: Reidel.
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.
Woodward, J. 2006. "Sensitive and Insensitive Causation." *Philosophical Review* 115: 1–50.
Ylikoski, P., and J. Kuorikoski. 2010. "Dissecting Explanatory Power." *Philosophical Studies* 148: 201–19.