TYPES OF CAUSES

Jeremy Freese

Northwestern University

J. Alex Kevern

Northwestern University

Prepared for *Handbook of Causal Analysis for Social Research*, Stephen L. Morgan (ed.).

ABSTRACT

The complexity of actual cause and effect relationships in social life can lead quickly to confused thinking and muddled discussions. Helpful here are distinctions that allow one to speak about some causes as different from others. Our essay describes several distinctions among causes that we find especially useful for social science. First, taking a broad view of what "causes" are, we discuss some issues concerning whether causes are manipulable or preventable. Then, we consider the distinction between proximal and distal causes, connecting these to concepts of of mediation and indirect effects. Next, we propose ways that concepts related the distinction between necessary and sufficient causes in case-oriented research may be also useful for quantitative research on large samples. Afterward, we discuss criteria for characterizing one cause as more important than another. Finally, we describe ultimate and fundamental causes, which do not concern the relationship between an explanatory variable and outcome so much as the causes of properties of the systems in which ordinary causal relationships exist.

The arrow salad: we have all seen examples before, and many of us have made them. One begins with the helpful convention of using boxes and arrows to specify consequential features of a system and the causal relations between them. But once one starts working theoretically—that is, trying to depict how we think some corner of human affairs *might actually work*—the boxes and arrows proliferate. Single-headed arrows turn into bidirectional ones as we contemplate feedback effects. Boxes are drawn within boxes and arrows collide with other arrows in our efforts to represent dimensionality and synergy.

The result is mess. We are left with both a vague sense that "everything causes everything else" and an anxious recognition that this revelation does little toward actually completing any research.

Social life comprises a series of nested complex systems, individual human organisms embedded in families, networks, and workplaces, which in turn are all embedded in nations and eras. The central trick of social inquiry is figuring out how to make orderly, accurate statements about these systems in the face of their enormous complexity and our limited capacity both to measure and to intervene. Social systems are dynamic, so outcomes of interest to one researcher figure as central causes of other outcomes studied by someone else. A voluminous literature considers the causes of differences in educational attainment, another voluminous literature considers the consequences of differences in educational attainment, and each of these is part of what motivates the other.

Social science explanations entail statements of cause and effect. But, as arrow salads illustrate, the actual cause-effect relationships in the world quickly overwhelm our

everyday sense of something as "a cause" of something else. Undisciplined thinking about causes leads quickly to muddled thinking, people talking past one another, unproductive rehashing of first principles, and confused data analysis. Meanwhile, the professional philosophical literature on casuality is often surprisingly unhelpful: the practical-minded research digs in looking for clarity and instead is soon invited to consider examples of simultaneous assassination attempts or billiard balls being rolled into time machines. No uncontroversial general philosophical account of causality exists, and social researchers have plenty of our own work to do while we wait.

What does exist are ways of expanding our working vocabulary of causality by means of useful distinctions among causes. Our essay is motivated by a conviction that one can avoid some of the confusion that results when thinking about complex, dynamic phenomena by better appreciating differences among causes--ways that not all the arrows in a box-and-arrows diagram are alike. In this essay, we articulate a few of the most important distinctions among causes that have been drawn in the social sciences and explain why we think these are especially useful to keep at hand.

COUNTERFACTUALS, PREDICTABILITY, AND MANIPULATION

Counterfactual dependence has become an essential part of the social science toolkit for thinking about causality. X is counterfactually dependent on Y if Y would be different had X been different. In the highly influential potential outcomes framework, one is invited to think about causal effects in terms of differences in the outcome (Y) over different states of an explanatory variable (X), even though only one state of X is observed for each case. For example, the effect of attending a private school for unit *i* is the difference

between achievement had *i* attended a private school and achievement had *i* attended a public school, when only one of these potential outcomes was actually observed and the defensibly estimating the other is the key matter of causal inference.

Equating causality with counterfactual dependence has important limits. As it turns out, philosophers can make short work of any simple version of idea (see essays in Collins, Hall, and Paul 2004). We will discuss this a bit more later, but part of the issue is that relations we recognize as causal combine two notions that typically, but not always, go together: causes *produce* effects, and effects *depend on* causes (Hall 2004). At the same time, it is important to keep in mind that the practical baseline for social science work is not advanced philosophy but rather commonplace intuitions. And grafting unsystematic commonsense intuitions about causality onto social science questions yields a mess. That this is so is evident, for example, by the need for social scientists to make clear when they intend to talk about *causal* effects, as if there is honestly any other kind of "effect." From the baseline, counterfactual thinking provides a significant cognitive upgrade for social scientists, even if later one may regret its limitations.

Counterfactual thinking provides especial clarity for how causal relations stand beyond association and beyond predictability. A nice foil here is provided by the concept of Granger causality in economics. As articulated by Granger (1969), X Granger causes Y if a time series of X is useful for predicting subsequent Y even after conditioning on preceding Y and preceding other variables Z, with Granger providing specific tests of this conditional association. The advantage is full explicitness about what one means when one asserts Granger causality. The key disadvantage is that Granger causality achieves this transparency by reducing causality to usefulness for prediction, and, depending on

what is outside one's data, X can Granger cause Y even though both are actually caused--in the sense of being counterfactually dependent--by something else. In other words, X can Granger cause Y even if X does not produce Y and independent changes in X would not yield changes in Y. From the counterfactual perspective, then, Granger causality is not a type of cause so much as not genuine causality at all, but rather an admirably well-operationalized form of conditional association.

Counterfactualist thinking invites thinking in terms of actual and hypothetical manipulations. This is congenial for social scientists, who are often interested in causes precisely so they can propose and evaluate interventions that would change outcomes. A temptation here is to take this a step further and make potential manipulability a criterion of causality. Holland touts a slogan "NO CAUSATION WITHOUT MANIPULATION" (capitalization in original, 1986: 959). More specifically, he recommends a distinction between causes and attributes, in which "causes are experiences that units undergo and not attributes that they possess" (2003: 8).

Holland (1986) cites the following as an example of the "confusion between attributes and causes" that he regards as pervasive in social science: "scholastic achievement affects the choice of secondary school" (p. 955). Holland argues that scholastic achievement is an attribute of the individual and not a cause of school choice because he cannot conceive of how scholastic achievement could be experimentally manipulated. Here, one might argue that this simply reflects a failure of experimentalist imagination on Holland's part (see also critique by Glymour (1986)).

More fundamentally, though, one can also imagine several distinct interventions that are consequential for school choice only through their effect on achievement,

so without the vocabulary of talking about achievement in a causal manner one can misapprehend how these interventions are actually working. Similarly, one may be interested in how different policies might affect school choice by changing the strength of the counterfactually dependent relationship between achievement and choice. In brief, events and attributes together provide the basic nodes of causal narratives--the verbs and adjectives, as it were, with units serving as the nouns--and outcomes may be counterfactually dependent on either. As Pearl (2009: 361) puts frankly, "Since Holland coined the phrase... many good ideas have been stifled or dismissed from causal analysis." (see also Bollen and Pearl, this volume).

Alongside Holland's distinction between causes and attributes, one also sees some make distinctions between "causes" and "enabling conditions," or between "events" and "conditions," where, in some sense, only events are asserted to be causal. In practice, our sense is this is typically not worth the argument, except in the need to be clear that one is not doing anything incoherent being interested in using a vocabulary of cause-and-effect to talk about how outcomes can depend on attributes or conditions and how contingent attributes and conditions can be involved in the production of outcomes. Being mindful of the distinction between events and attributes or conditions is valuable when constructing explanations, but restricting the use of "cause" to the former seems to us overly restrictive in practice.

Of course, whether a cause can actually be manipulated is valuable for assessing the potential for intervention. Epidemiology uses the concept of "preventable" causes of death to refer to those causes that can be modified by behavioral changes or relatively straightforward interventions. For example, Danaei et al. (2009) identifies the four leading

preventable causes of death as smoking (~19% of all 2005 deaths), high blood pressure (~16%), overweight/obesity (~9), and physical inactivity (~8). Key to note among these preventable causes of death is precisely that they are not the causes of death that are listed as primary causes of death on a death certificate or on the leading causes of death list (e.g., isochemical heart disease, cancer, and stroke). Rather, each of these preventable causes of death is a manipulable behavior or attribute that is linked to multiple, more proximate causes of death. In other words, whether and how causes can be manipulated is vital for policy, but recognizing this does not require stipulating manipulability as a criterion of what can be called a cause.

CASUAL PROXIMITY

We can illustrate the basic issue of causal proximity with an example from Leahey (2007), who begins by noting that, on average, female academics in many fields are paid less than men and also have lower research productivity (Leahey 2007). Imagine if someone were to argue that the entire reason female academics are paid less than men is that they are less productive. We might then diagram the proposed relationship between gender (*G*), productivity (*P*), and salary (*S*) like this:

In this case, setting aside the above concern about attributes vs. causes, both gender and productivity may be properly characterized as causes of school achievement, but productivity is a more proximal cause and gender is a more distal cause.[1] In other words,

[1] There is a specific sense to the legal use of the term "proximate cause" that we leave outside the scope of this essay.

the cause of a cause of the outcome is still a cause, just a more distal one.

Distance here is a matter of the length of the chain of more proximate causes mediating the relationship between a cause and outcome. This is not some natural fact about the world but a matter of the level and kind of causes we are considering. For instance, Leahey proposes the relationship between gender causes differences in the degree of specialization by academics, and degree of specialization (*S*) enhances productivity. She also posits that productivity differences cause differences in visibility (*V*) among academics, and differences in visibility cause differences in salary. The best-fitting model in her analysis ends up as:

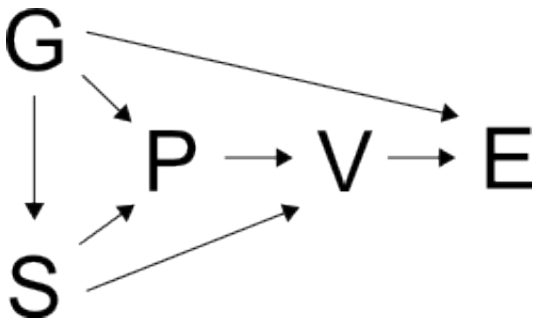
Productivity is now a distal cause relative to the newly added construct of visibility, and gender is now several variables distal from salary. We have not changed anything about the natural world in the move from the first diagram to this one; what we have done is potentially elaborate our theoretical understanding.

Indeed, progress in social science often proceeds precisely by establishing intervening relationships that make a previously inadequately understood causal relationship more distal. This is perhaps especially so in the study of social inequalities, as typically there the animating questions are not "what causes Y?" by "why do groups X differ in Y?," thereby putting the questions of whether X is really a cause of Y and why at the fore. Competing theories of why X causes Y often turn on different implications about more proximate causes. An important criticism of the increased emphasis on "natural experiments" and instrumental variables techniques in causal inference is precisely that

these techniques often offer little or no leverage for analyzing mediating relationships (Morgan and Winship 2007).

In graphs (a) and (b), proximate causes are depicted as *strictly mediating* more distal causes. Cause Z strictly mediates the causal relationship between X and Y if the causal relationship is exclusively due to X causing Z and Z causing Y. In terms of interventions, strict mediation means that if we were able to intervene and stop the causal influence of either X on Z or Z on Y (for example by equalizing Z on all cases), then interventions on X would no longer affect Y. In other words, if a hypothetical intervention eliminated gender differences in specialization, then we would no longer expect to observe gender differences in productivity, visibility, or salary.

More commonly, social science proceeds by identifying more proximate causes that partially, rather than strictly, mediate the relationship between a cause and an outcome. For example, the theoretical model that Leahey proposes for the relationship between gender and salary is actually:



In this model, Leahey proposes that gender differences in specialization are part, but not all, of why gender is causally related to productivity. Also, gender differences in productivity are part, but not all, of why gender is causally related to salary differences.

The hypothetical intervention that eliminated in gender differences in specialization may be thus expected to reduce, but not eliminate, gender differences in salary.

The most common language for talking about partially mediating relationships is that of direct, indirect, and total causal effects. Leahey's theoretical model also proposes that productivity partially, but not strictly, mediates the relationship between specialization and visibility. The total causal effect of specialization on visibility corresponds to the change in visibility which results from a change in degree of specialization, regardless of the mechanisms involved. The total causal effect can then be decomposed into indirect effect(s) via specific, partially mediating, proximate cause(s), and the remaining direct effect.

The indirect effect of specialization on visibility here is the effect changes in specialization would have if the only way specialization affected visibility was through productivity. This is a tricky concept for the counterfactual frameworks, and Pearl (2009: 132) calls the indirect effect "a concept shrouded in controversy and mystery." His solution is to consider indirect effects a quantity that depends on two separate counterfactuals. First, one estimates what values of mediator Z we would have observed under counterfactual values of X. Then, we estimate what values of outcome Y we would have observed if X was held to its actual values but Z was changed to their estimated values from the first counterfactual.

The direct effect of specialization on visibility is simpler: it is the effect changes in specialization would have if those changes were somehow blocked from having any influence on productivity. An easy way to conceptualize this in counterfactual terms is to imagine a joint intervention in which values of X are changed but Z is artificially held

constant.

What is crucial to keep in mind in such analyses is that "direct" effects here are only direct given the variable(s) for which indirect effects are being estimated. An estimated direct effect may be entirely mediated by more proximate causes not in the model. In other words, a direct effect is a residual finding about how much of a distal causal relationship remains unaccounted for after specific more proximal causes are considered. It does not imply anything further about the immediacy of the process by which the cause brings about the outcome.

Note also that when we acknowledge that the causes of causes of an outcome are themselves causes, we acknowledge that the number of causes of an outcome is indefinite, akin to how our number of ancestors is indefinite and, if we go far enough back, may include most everyone alive at the time. This makes some people suspicious. For example, Martin (2011: 38) presents two scenarios: (1) A sells B a store and a year later, C breaks into the store and kills B and (2) in which A was an Pleistocene-era ancestor of C, who kills B. In each case, he asks "Did A cause the death of B?" and his answer of "yes" leads him to conclude that, in the counterfactualist framework, "we cannot ask the question, 'What caused B's death?' and bring in anything less than an infinite number of causes, with little way of telling them apart."

Martin is correct that once one gets into (causes of) causes of causes, historical questions like 'What caused B's death?' or 'What caused World War I?' do not have a clear stopping point. Where his reasoning errs is in its suggestion that we are powerless to draw useful distinctions among causes nevertheless. From the above, for

example, we can distinguish that C murdering B is a more proximate cause of B's death than A's selling him a store. In our discussion of causal importance below, we can likewise determine that C murdering B is a more important cause of B's death. At the population level, these same pseudo-conundrums either lead to causes that are so indirect that their influence on the outcome is beneath whatever threshold of trivial (Martin's shop-seller example), or causes that pertain to explaining the existence of the population rather than variation in the population (his Pleistocene example).[2] In other words: yes, anybody who has watched enough time-travel movies appreciates how any single event is the culmination of a whole plenum of things that could have happened differently, but this has no crippling implication for the use of counterfactuals as the major conceptual workhorse for thinking about causality and causal explanation in social science.

CAUSAL CONFIGURATIONS

Perhaps the major divide in causal analysis in social science separates "case-oriented" and "population-oriented" (or "variable-oriented") endeavors. Case-oriented projects "seek to explain particular outcomes in specific cases" (Mahoney 2008: 414). Many such projects are nevertheless comparative and seek to make general statements that apply across multiple cases. An example of a comparative case-oriented question would be "When do austerity programs result in severe social protests?" (see Ragin 2000). Population-oriented projects seek to make general statements about the distribution of causal effects over populations or subgroups of populations. An example of population-

[2] By "population level" here, we mean statements intended to apply to multiple cases rather than statements about the causes of an outcome for a single case. This is sometimes referred to as the distinction between singular causes and general causes (e.g., Pearl 2009: 253-256).

oriented question would be "Does growing up in a bad neighborhood affect school achievement? If so, how much is this effect, why does it exist, and does it vary in systematic ways across persons?"

This essay, like the rest of this volume, is predominantly concerned with causal distinctions as they pertain to the practice of population-oriented social science. That said, case-oriented researchers has made a vigorous effort in recent decades at articulating the logic of causal inference from comparative case-study data, especially in terms of establishing the limitations of thinking of causal inference for such data in statistical terms (see, e.g., essays in Brady and Collier 2010). Moreover, one way that counterfactual frameworks are cognitively useful for population-oriented research is that they heighten attention to the fact that the effects estimated by regression-type models, when causal, are summaries of case-level causal effects. Populations are comprised of individual cases even if the researcher is only interested in aggregate statements. As such, population-oriented approaches need to be compatible with the explanation of individual cases.

Two fundamental logical distinctions that are common in case-oriented research but practically absent in large N variable-oriented research are **necessary causes** and **sufficient causes**. Saying that X is a necessary cause of Y implies that some state of X is needed in order for some value of Y to occur. To say, for instance, that contracting HIV is a necessary cause of developing AIDS is to imply that nobody has AIDS who does not first have HIV. Saying that X is a sufficient cause of Y implies that some state of Y will occur if some state of X occurs. Prior to medical developments, rabies was a sufficient cause of mortality; every single person who contracted rabies died fairly shortly thereafter.

For a binary cause and outcome, both cases imply an empty cell in a 2x2 table (no HIV-free AIDS victims; no rabies survivors). Rarely when you do large-N population-based research do you actually see an empty cell that is not based on some mechanical aspect of the data. There are various reasons that necessary or sufficient conditions observed in case-oriented research may be far more elusive in population-oriented research, but two stand out. The first reason is that observing necessary and sufficient causes demands accurate operationalization and measurement. In survey research, for example, large samples often contain enough measurement errors that even logically mandatory relationships will often not appear as such in survey research unless specifically imposed by investigators. Somebody reports having never attended college and yet being employed as a physician; somebody else reports five sexual partners in one wave of a survey and being a virgin in the next. The second reason is that since populations encompass many more cases, they are much more likely to include genuinely exceptional cases that negate the necessary or sufficient causal claim. Dion (1998) suggests that "probabilistically necessary" and "probabilistically sufficient" can be used for large N where either measurement issues or the possibility of unaccountably idiosyncratic processes lead to relations that are still useful to talk about in quasi-deterministic terms even though observed exceptions exist.[3]

What may be necessary or sufficient to produce an outcome is not a single cause but rather a configuration of causes. Key concepts here are INUS-causes and SUIN-

[3] A key philosophical issue that recurs in discussing the relationship between case- and population-oriented approaches concerns the extent to which outcomes for individual cases are truly probabilistic versus the apparently probabilism simply reflecting inadequate information (Mahoney 2008, Lieberson 1991).

causes. A **SUIN cause** is a Sufficient but Unnecessary part of a causal condition that is itself Insufficient but Necessary. Mahoney (2008) gives the example of the democratic peace theory, in which the absence of democracy is necessary (but not sufficient) for war. If different conditions are sufficient to undermine democracy (e.g., "fraudulent elections," "repression"), then these conditions are SUIN causes of war. By undermining democracy, fraudulent elections are neither necessary nor sufficient for war, but they do enable the possibility.

A general example of SUIN causes may be **precipitating causes**, events that bring about an outcome in the presence of other, enabling conditions. Riots generally follow a preceding event (like the Rodney King beating verdict and the Los Angeles riots of 1994). Yet, that event is typically understood as not a sufficient cause--deteriorating conditions made the area in question ripe for a riot--or a necessary cause (other events, had they happened instead, could have triggered a riot given the same conditions). The occurrence of some precipitating event may therefore be a necessary but not sufficient condition for a riot, and any of a number of possible events may be sufficient but not necessary to serve as a precipitating event.

INUS causes have received more attention. An INUS-cause is an insufficient but necessary part of a causal condition that is itself unnecessary but sufficient (Mackie 1965; sometimes "non-redundant" is used instead of necessary here). For example, in studying when countries undertake policy reform, rightist partisanship has been identified as an INUS cause of unpopular reform. If so, the existence of rightist partisanship does not itself bring about reform, and unpopular reform can occur in the absence of rightist partisanship, but there are specific conditions under which unpopular reform will happen in the

presence of rightist partisanship but not in its absence.

INUS causes are known as **component causes** in epidemiology (Rothman and Greenland 2005; Johansson and Lynøe 2008). The idea here is that there are many different configurations of causes that are each sufficient to produce a disease. A component cause influences the outcome by being part of at least some of these configurations. Having unprotected sexual intercourse is neither necessary nor sufficient to contract HIV, yet of course many people who are HIV positive who would not be had they never had unprotected sex. One possibility conceptualization then is that there are various conditions sufficient for HIV transmission to occur, and unprotected sexual intercourse is a necessary part of some but not all of these conditions.

Population-oriented research typically works with imperfect measures on only a subset of the actual causes that influence outcomes. When an outcome is produced by the realization of one configuration of a large set of sufficient causal configuration, and some INUS causes important to the configuration are unobserved, then the outcome will look like it has a probabilistic relationship to the INUS causes that we do observe. In other words, a deterministic world replete with INUS causality is consistent with a world that can only be modeled in statistical terms with largely additive causal relationships when the variables in the model comprise only a modest subset of all relevant causes (Rothman and Greenland 2005 and Mahoney 2008 provide a nice juxtaposition of this point from the perspectives of epidemiology and comparative-historical social science, respectively). There has been some work in epidemiology on a sufficient component cause model that conceptualizes component causes as such rather than as additive and interacting terms in a conventional model, but this work has not yet reached the point where its applicability to practical social

science research problems has been demonstrated (e.g., Flanders 2006, VanderWeele and Robins 2007). Nevertheless, the idea that additive models may be estimating what are really highly complex and possibly deterministic component cause relationships allows one way of connecting the logic of population models with case-specific explanations.

CAUSAL IMPORTANCE

In their excellent chapter on "causal asymmetries," Wright, Levine, and Sober (1992) cover a variety of rationales by which it could be asserted that one cause is "more important" than another. Obviously, such assertions might serve a variety of rhetorical purposes. Their position, in the end, is that various kind of *qualitative* distinctions that one can make about causes do not provide systematic grounds for declaring one cause more important than another. Rather, the only grounds on which they conclude that claims about relative importance of causes to be consistently, coherently made is if they can be articulated and adjudicated in quantitative terms.

In regression models that afford causal interpretations, coefficients can be interpreted as the effect of a unit increase of x on y. Comparing the magnitude of two coefficients to determine which is the most important cause raises the obvious problem that the magnitude of coefficients depends on what scale we choose: we could make age an arbitrarily more or less important cause of an outcome by changing measurement from seconds to centuries. A common approach is to allow the population distribution of our variables to define what comprises a commensurate metric for us by using the population distribution. The prime example here are x- of *fully*-standardized coefficients--that is, regression coefficients based on measuring x in standard deviations.

Of course, one typically does not know the population distribution of X, but must estimate it using the observed distribution of X in one's sample. For standardized coefficients to make any sense whatsoever, the standard deviations on which they are based must be meaningful quantities in terms of the population whose parameters we are attempting to estimate. This is important to keep in mind because, in properly specified models, sample-based regression coefficients do not actually have to be based on representative samples in terms of X--weighted or unweighted--in order for coefficients to be unbiased, but the same cannot be said for standardized regression coefficients.

Beyond this, many commentators have been critical of comparing standardized coefficients to evaluate relative importance. If we are talking about comparing coefficients in a single model, one may note the counsel of Winship and Sobel (2004: 499) that successfully estimating the effect of one independent variable on an outcome is sufficiently difficult that "attempts to estimate the effects of multiple variables simultaneously are generally ill advised." Blalock (1961: 868) more specifically advises that it is "unwise to become involved with the problem of evaluating the relative importance of variables which stand in some sort of causal relationship to each other"; that is, to compare the importance of distal and proximal causes. If one is to forge ahead, the claims about the relative importance of two causes would seem to imply a comparison of the *total* causal effects of each variable, which would imply either a structural equation modelling approach or one based on estimating the total causal effect of the distal and proximate cause as separate models.

A different, more questionable, idea is that comparing standardized coefficients is

misguided because it conflates the estimated effect with the variance of *X*, when typically researchers are only interested in the former (King 1986: 671). The same principle, that standardized coefficients are problematic because of their dependence on the variance of *X*, is also sometimes used to argue that comparing standardized coefficients is especially a bad idea when comparing coefficients across groups (Treiman 2009). To consider an applied example, Branigan et al. (2011) found that skin color measured in objective terms (percent reflectance) has about the same estimated unstandardized coefficient with educational attainment for white men and for black men. For purposes here, let us presume in both cases the coefficient indeed does accurately estimate the total causal effect of skin color differences on educational attainment (via, e.g., differential treatment by teachers or peers).

Does this imply that skin color is an equally important cause of educational differences for black men and white men? Variation in skin color reflectance is much larger for blacks than for whites. So while the unstandardized coefficient is the same, the standardized coefficients are much different; if you look at the difference between the 25th and 75th percentiles for each group, the expected education difference for black men is twice as large as it is for white men. To us, this implies that skin color is more important cause of educational attainment for black men than white men even though the unstandardized coefficients are the same (see also Hargens 1976), while others have drawn the opposite conclusion from comparable worked examples (Treiman 2009: 110).

More broadly, we think the question of "Does X_1 or X_2 have a bigger effect on Y?" differs from "Is X_1 a more important cause of X_2 than Y?" precisely in that the latter question depends in part on how common the two causes are. Consider again the study that found

that smoking was the leading preventable cause of death in the United States. The sense here by which smoking is said to be "leading" is that smoking is estimated as having killed the most people of any cause in the set of preventable causes. We regard this as a defensible warrant for saying that it is the most important of these causes. But this is not directly a claim about the magnitude of the size of the effect of smoking on the mortality prospects of individuals--many behaviors are more lethal than smoking for those who engage in them, but not as many people engage in them. That is, the number of people killed by smoking is a function of both (1) how lethal smoking is for those who smoke and (2) how many people smoke.

Epidemiologists refer to this sometimes as the *population attributable fraction*. There are complexities here depending on particularities of the application (see Greenland and Robins 1988), but a rough way of thinking about this quantity in counterfactual terms is to consider the difference between the actual population distribution of the outcome and the potential distribution if X was held constant across all cases (e.g. if no one smoked). A simple expression of this quantity can be computed as $p(X)[p(Y|X) - p(Y|\sim X)]$.

Importance here is a population-specific determination, in that it depends on the particular distribution of causes in the population, as well as on the distribution of anything else that would cause the effect size to differ over individuals in the population. Populations can be divided into strata (e.g., age groups) and effects estimated within strata; these can be used to evaluate how the relative importance of different causes changes with actual or hypothetical changes in the population distribution (Greenland and Rothman 2008). Of course, the magnitude of effects themselves may change as populations change. For example, there has long been debate about how the value of an educational credential

for wages changes as the percentage of the population with that credential increases.

Also, outcomes are not exclusively attributed to particular causes; that is, the sum of attributable fractions across all cases is greater than 1 (Rothman and Greenland 2005). Consider the example of causes that operate synergistically, as in if smoking was much more lethal among obese people than nonobese people. Then, there would be overlap in the counterfactual survivors of an intervention that prevented anybody from smoking and an alternative intervention that prevented anybody from becoming obese.

The same could be said for causes that stand in a distal/proximate relationship to one another. Presume that low physical activity is a cause of obesity and vice versa. Again, this implies overlap in the counterfactual survivors of an intervention that increased physical activity (and reduced obesity indirectly) and an alternative one that reduced obesity (and increased physical activity indirectly).

To give another example, in behavioral genetics, variance decomposition techniques are often used toward generating findings about the relative importance of genetic variation versus environmental variation, as in, for example, a report that genetic variation is more important than environmental variation in determining height. Yet part of how genes can influence outcomes is by influencing traits that influence the experience of environmental exposures--children who evince an early aptitude for reading report enjoying reading more, are encouraged to read more, and spend more time reading (Rutter 2006). For the variance decompositions of behavioral genetics to add to 100 percent, one either needs to assume that there are no interactions or correlations between genes and environments, or that the decomposition is actually between genes and exogenous environments, that is, environmental influences that are independent of genetic

endowments (Freese 2008).

Typically, whether one cause is more distal or proximate than another does not indicate whether it is more important in a quantitative sense. Again, the relative size of the total causal effect would seem to be at issue, and either a distal or proximate cause can have a larger total causal effect.[4] In the health disparities literature, there is a longstanding debate about the relative benefit of "upstream" (distal) interventions versus "downstream" (proximate) ones, which turns on downstream interventions having bigger effects on particular pathways to disease but upstream interventions potentially exerting influence through many more pathways.

For that matter, we should underscore the crucial practical difference between changes in an outcome under a hypothetical equalizing intervention and the changes that may be anticipated by actually available interventions. If one can actually intervene more on A than on B--either in absolute terms or in terms of what can be attained for the same cost--then intervening on A instead of B may have a greater effect on the full distribution of Y even if B is a more important cause in the sense of the attributable fraction.

For the kinds of causes that are prominent within case-oriented research, Mahoney, Kimball, and Koivu (2009) offers a technique based on Venn diagrams that depict the sets of cases in which the cause and outcome occur. For a necessary cause, the set of cases in which the outcome occurs is entirely subsumed within the set of cases in which the cause occurs. The opposite is true for sufficient causes: the set of cases in which the cause

[4] The exceptions are if the distal cause entirely determines the more proximate cause or if the distal cause is strictly mediated by the more proximate cause. In the former scenario, the total causal effect of the distal cause must be at least as large as the total effect of the proximate cause, whereas in the latter scenario the reverse is true.

occurs in entirely subsumed in the set with the outcome. The relative importance of two necessary or two sufficient causes, then, may be adjudged by which is closer to being a necessary and sufficient cause, in which case the two circles would be exactly coterminous. Equivalently: of two necessary causes, the more important cause is the less common one; of two sufficient causes, the more important cause is the more common one.

For SUIN and INUS causes, Mahoney, Kimball, and Koivu (2009) suggest that these causes are more important the extent to which they approximate necessary (for SUIN) or sufficient (INUS) causes. This is more debatable. The implication is that any sufficient cause is more important than any INUS cause. The tension, analogous to our discussion of quantitative comparison, comes in comparing a rare sufficient cause to a common INUS cause. A very rare radiation exposure that is sufficient for developing lung cancer is still hard to consider as important a cause of lung cancer as smoking, even though smoking is not only an insufficient cause of lung cancer, but most smokers do not get lung cancer (example adapted from Wright, Levine, and Sober 1992). Again, the issue is that, because smoking is so common, many more cases of lung cancer are attributable to it than to the radiation exposure.

A better standard, a one that makes more consistent use of the Venn diagram technique, may be simply to judge one INUS cause as more important than another INUS cause to the extent to which it approaches a necessary and sufficient cause--that is, to the extent to which the cause and outcome are coterminous. If we use X to indicate the presence of the cause and Y to indicate the presence of the outcome, this can be expressed as $p(X,Y)/[p(X)+p(Y)-p(X,Y)]$.

Note that there is a slight difference here between the conceptualization that

underlies this formulation and that which underlies the attributable fraction. When answering the question of how many deaths are attributable to smoking, the number of smokers who do not die is irrelevant. By that standard, if smoking and obesity killed the same number of people each year in terms of the attributable fraction, we would consider them equally important causes even if obesity was more common than smoking. In terms of the Venn diagram technique, however, if smoking and obesity killed the same number of people, this would mean they had the same overlap with the cause (that is, that they were equally sufficient causes), but because smoking was rarer it would have less area outside the cause. We would therefore judge smoking to be a more important cause than obesity because it was closer to being a necessary cause. The broader point is that either standard provides a consistent and coherent way of determining which of two causes is more important, but what differs is whether what we consider important is accounting for the occurrence of a binary outcome (which is what the attributable fraction does) or distinguishing occurrences from non-occurrences.

CAUSES OF CAUSAL RELATIONSHIPS

Lieberson (1985) distinguishes between **surface causes** which "appear to be generating a given outcome" and **basic causes** which "actually generate an outcome" (p. 185). The hypothetical example he provides is of a gap in income between racial groups that appears attributable to educational differences, but reductions in the educational differences do not lead to any change in the income gap. Lieberson's example permits multiple interpretations, and a trivializing one would be to say just that surface causes are not properly causes at all, but simply exemplify spurious association between a real ("basic")

cause and the outcome.

Two more interesting possibilities, however, each call attention to distinctions arising from how simple estimates of causal effects can be misleading given broader dynamics of the system in which they occur. The first possibility is that the basic cause and outcome could be linked by a number of different surface causes in such a way that what is actually generating group differences in a given context is effectively redundant with other potential causes of group differences. Earlier, we mentioned that causality encompasses two notions that are not entirely the same: that causes produce outcomes and that outcomes depend on causes.

Redundancy in causal systems provides one case in which the divergence of these two notions may be clear. In a given case, for instance, an educational difference may provide the grounds on which a minority-race candidate is passed over for a job in favor of a majority candidate. When educational credentials are equal, however, perhaps other characteristics that would have earlier disadvantaged minority-race candidates with less education come to the fore (like perceived fit with existing employees), which lead again to the minority candidate being passed over. In other words, by tracing the causal process in given cases (e.g., Bennett 2010), we might come to proper inferences about causes that produced the outcome in those cases that nevertheless overstate the changes that would result from intervening on the cause.

Many information systems, like telecommunications systems, are designed to be redundant--the system will still convey messages from A to B even if a node that is normally part of the actual connection used to link A and B is removed. In the philosophical literature, one popular toy example involves someone who is fatally shot

after having ingested a poison that would have killed them otherwise--one can say the shooting was the **actual cause** of death and yet the outcome would have been the same had it not happened. Lieberson's distinction between basic and surface causes may suggest the analogous possibility at the level of social dynamics: that an mediating variable like education may account for racial differences in one context, but that disadvantages are sufficiently redundant that interventions on education do not actually affect the ultimate magnitude of the race gap.

An alternative possibility suggested by Lieberson's example is that the basic cause and outcome could be linked by a mechanism implicated in the generation of the surface causes themselves, such that addressing one surface cause leads to another surface cause emerging or increasing in importance. Consider a democratic society that includes one region in which members of the dominant ethnic group wish to dilute the voting power of a subordinate group. A literacy test is instituted that accomplishes this purpose. Egalitarian-minded courts ban the use of these tests. The dominant group responds by instituting a poll tax, which has much the same effect on participation by the subordinate group that the literacy test did.

In this example, when the literacy test was in place, it was the surface cause of group differences in electoral participation in the sense that it served as the proximate means whereby the group difference was produced. Yet the difference in participation was not counterfactually dependent on the existence of a literacy test so long as the more basic cause existed of the dominant group wishing to suppress participation by the subordinate group and having various other available means of doing so.

In Lieberson's example of educational differences and income differences, imagine

if the basic cause of income differences was employers being strongly disinclined to hire minority-race workers. Educational differences may then provide a pretext for not hiring black workers, but, if education were equalized, employers would emphasize some other criterion that disadvantaged black workers. The difference between this example and the earlier example of causal redundancy is that here part of why the proximate causal relations exist and are sustained is their role in preserving the relation between a distal cause and outcome.

These more systemic interpretations of basic and superficial causes presage the concept of **fundamental cause** that has become a central concept of epidemiological sociology (Phelan, Link, and Tehranifar 2010; Link and Phelan 1995). The concept has been used primarily as a potential characterization of the inverse relationship between socioeconomic status and health. Its usage is more easily understood against a backdrop in which some have regarded SES as simply a placeholder construct to be supplanted by "real" causes of population health, or regarded SES as a real cause but too distal to be of value for epidemiology beyond highlighting an ignorance to be resolved by a search for mediating variables. The problem with the backdrop view is that the relationship between socioeconomic status and health has largely proven more robust that the more immediate causes of disease and death that prevail in a particular population at a particular time. Roughly: what kills people changes, but that lower status members of society die earlier does not.

Lutfey and Freese (2005) argue that fundamental causality is thus a distinct logical type of causal relationship. For X to be a fundamental cause of Y, X must be a distal cause with many proximate consequences, and Y must be an outcome with many proximate

causes. Consequently, X and Y are typically linked by massively multiple mechanisms, and a systematic asymmetry exists among these mechanisms such that those by which X influence Y in one direction are much greater in number and magnitude than the mechanisms by which X influences Y in the other. In other words, the detailed pathways by which low social standing may negatively influence health are vast, and overwhelming in comparison to the ways that low social standing positively influences health. Then, there must exist some meta-mechanism or durable mechanism that accounts for the preservation of this asymmetry as mechanisms change.

Link and Phelan emphasize "flexible resources" as a meta-mechanism linking SES and health: good health is a broadly desirable end and socioeconomic status provides differential means in achieving that end. Freese and Lutfey (2011) distinguish the claim that SES is a fundamental cause of health from any particular theory of the durable narrative involved, and they raise the possibility of spillovers as a durable narrative separate from "flexible resources" that may be important for understanding enduring and robust health disparities. Regardless, note that SES as a fundamental cause is not an academic claim devoid of policy implications: the implication is that differences in social standing and the capacity to use means to protect health are together sufficient for health disparities. In other words, calls to "eliminate" health disparities without addressing resource differences are likely fanciful, and the real effect of interventions on disparities may depend on their overall effect on the capacity for agentic behavior to protect health.

Relatedly, evolutionary biologists and psychologists sometimes distinguish proximate and **ultimate causes** (Mayr 1961; Laland et al. 2011). Consider the theory that father absence influences the pace of pubertal development because, in our species history,

father absence provides a proxy for the amount of a paternal investment a woman's own children would receive, and the optimal pace of development in terms of reproductive fitness is accelerated in low-investment versus high-investment environments (Belsky, Steinberg, and Draper 1991). (Set aside whether this theory is actually true.) In this theory, father absence is a proximate cause of differences in pubertal development. The implication is that we would expect manipulations of father absence would lead to differences in development and that some mechanism exists linking immediate consequences of father absence to the immediate physiological causes of different rates of development.

But "ultimate" causes here are not the same as the "distal" causes discussed earlier, even though both terms were contrasted with proximate causes. Distal causes in this example would be causes of father absence. Ultimate cause, on the other hand, makes reference to the possibility of a historical explanation for the development of the embodied physiological mechanism responsible for the causal relationship between father absence and pubertal development. That claim entails either direct historical information or some theory of the "logic of history." In this case, the logic of history is provided by the shaping of physiology over generations by evolution by natural selection, and the theory is that the fitness advantages associated with an adaptive timing of pubertal development caused physiological mechanisms responsive to father absence to evolve in our species history.

Ultimate causes do not have to reference species history or natural selection. Functionalist explanations trace the origins and sustenance of causal relationships to larger systemic imperatives. A classic example here is Malinowski's explaining the elaborate fishing rituals of the Trobriand Islanders by their effects on reducing fears associated with

an intrinsically dangerous task (Stinchcombe 1968; Wright, Levine, and Sober 1992). The implication is that a counterfactualist who came ashore with Malinowski would observe the Islanders and come away with a causal story about the fearfulness reducing effects of the ritual. While correct, this would miss a vital part of the phenomenon, which is the role of this causal relationship in explaining why the Islanders conduct the ritual in the first place. If Islanders were prevented by outsiders from observing this ritual--but not from fishing!--we might expect the development of some alternative cultural or institutional mechanism for reducing fear. Likewise, if changes resulted in fishing no longer being as otherwise fear-provoking, the rituals may persist by cultural inertia but would not have the same dynamic resisting their discontinuation or evolution to a different form.

CONCLUSION

Societies are enormously complex systems and so social science is an extraordinarily complex project. A temptation toward making the enterprise more tractable is to focus on narrow questions of assessing interventions. While obviously important, there are many puzzles to social life that cannot be reduced to analogies of program evaluation. Even so--or, especially so--questions about complex causal relationships in social research require clear and disciplined thinking about the structure of causal relationships if they are to be successfully engaged. In this essay, we have focused on distinctions that can be made among causes and have tried to explicate aspects of several of the most handy ones. To be sure, not every complexity of social science explanation can be reduced to finding just the right adjective to put in front of "cause," but recognizing how fundamentally different kinds of causes can be complements toward a more complete understanding of a phenomenon

provides useful cognitive tools.

REFERENCES

Belsky, Jay, Lawrence Steinberg, and Patricia Draper. 1991. "Childhood experience, interpersonal development, and reproductive strategy: An evolutionary theory of socialization." Child Development 62:647-670.

Bennett, Andrew. 2010. "Process Tracing and Causal Inference." in Rethinking Social Inquiry: Diverse Tools, Shared Standards (Second Edition), edited by H. E. Brady and D. Collier. Lanham, MD: Rowman & Littlefield.

Brady, Henry E. and Andrew Collier. 2010. Rethinking Social Inquiry: Diverse Tools, Shared Standards, Second Edition. Lanham, MD: Rowman and Littlefield.

Danaei, Goodarz, Eric L. Ding, Dariush Mozaffarian, Ben Taylor, Jurgen Rehm, Christopher J. L. Murray, and Majid Ezzati. 2009. "The preventable causes of death in the United States: comparative risk assessment of dietary, lifestyle, and metabolic risk factors." PLoS Medicine 6:e1000058.

Dion, Douglas. 1998. "Evidence and Inference in the Comparative Case Study." Comparative Politics 30:127-145.

Flanders, Dana. 2006. "On the relationship of sufficient component cause models with potential outcome (counterfactual) models." European Journal of Epidemiology 21:847-853.

Freese, Jeremy and Karen E. Lutfey. 2011. "Fundamental causality: Challenges of an animating concept for medical sociology." in Handbook of Medical Sociology, edited by B. Pescosolido, J. Martin, J. McLeod, and A. Rogers. New York: Springer.

Granger, Clive W. J. 1969. "Investigating causal relations by econometric models and cross-spectral models." Econometrica 37:424-438.

Greenland, Sander and James Robins. 1988. "Conceptual problems in the definition and interpretation of attributable fractions." American Journal of Epidemiology 128:1185-1197.

Greenland, Sander and Kenneth J. Rothman. 2008. "Introduction to Stratified Analysis." Pp. 258-283 in Modern Epidemiology, 3rd Edition, edited by K. J. Rothman, S. Greenland, and T. L. Lash. Philadelphia, PA: Lippincott, Williams, and Wilkins.

Hall, Ned. 2004. "Two Concepts of Causation." in Causation and Counterfactuals, edited by J. Collins, N. Hall, and L. A. Paul. Cambridge, MA: MIT Press.

Hargens, Lowell L. 1976. "A Note on Standardized Coefficients as Structural Parameters." Sociological Methods and Research 5:247-256.

Holland, Paul. 1986. "Statistics and causal inference." Journal of the American Statistical Association 81:945-60.

Holland, Paul W. 2003. "Causation and Race." Educational Testing Service Research Report RR-03-03.

King, Gary. 1986. "How Not to Lie With Statistics: Avoiding Common Mistakes in Quantitative Political Science." American Journal of Political Science 30:666-687.

Laland, Kevin N., Kim Sterelny, John Odling-Smee, William Hoppitt, and Tobias Uller. 2011. "Cause and effect in biology revisited: Is Mayr's proximate-ultimate distinction still useful?" Science 334:1512-1516.

Leahey, Erin. 2007. "Not by productivity alone: How visibility and specialization contribute to academic earnings." American Sociological Review 72:533-561.

Lieberson, Stanley. 1985. Making It Count: The Improvement of Social Research and Theory. Berkeley and Los Angeles: University of California Press.

Lieberson, Stanley. 1991. "Small N's and Big Conclusions: An Examination of the Reasoning in Comparative Studies Based on a Small Number of Cases." Social Forces 70:307-320.

Link, Bruce G. and Jo C. Phelan. 1995. "Social Conditions as Fundamental Causes of Disease." Journal of Health and Social Behavior 35:80-94.

Lutfey, Karen and Jeremy Freese. 2005. "Toward some fundamentals of fundamental causality: Socioeconomic status and health in the routine clinic visit for diabetes." American Journal of Sociology 110:1326-1372.

Mackie, J. L. 1965. "Causes and Conditions." American Philosophical Quarterly 2:245-264.

Mahoney, James. 2008. "Toward a unified theory of causality." Comparative Political Studies 41:412-436.

Mahoney, James, Erin Kimball, and Kendra L. Koivu. 2009. "The Logic of Historical Explanation in the Social Sciences." Comparative Political Studies 42:114-146.

Martin, John Levi. 2011. The Explanation of Social Action. Oxford: Oxford University Press.

Mayr, Ernst. 1961. "Cause and Effect in Biology." Science 134:1501-1506.

Pearl, Judea. 2009. Causality: Models, Reasoning, and Inference (Second Edition). Cambridge: Cambridge University Press.

Petersen, Maya L., Sandra E. Sinisi, and Mark J. van der Laan. 2006. "Estimation of Direct Causal Effects." Epidemiology 17:276-284.

Ragin, Charles C. 2000. *Fuzzy-Set Social Science*. Chicago: University of Chicago Press.

Rothman, Kenneth J. and Sander Greenland. 2005. "Causation and causal inference in epidemiology." American Journal of Public Health 95:S144-S150.

Rubin, Donald B. 2004. "Direct and Indirect Causal Effects via Potential Outcomes." Scandanavian Journal of Statistics 31:161-170.

Rutter, Michael. 2006. Genes and Behavior: Nature-Nurture Interplay Explained. Malden, MA: Blackwell.

Stinchcombe, Arthur. 1968. Constructing Social Theories. New York: Harcourt, Brace and World.

Treiman, Donald J. 2009. Quantitative Data Analysis: Doing Social Research to Test Ideas. San Francisco, CA: Jossey-Bass.

VanderWeele, Tyler and James M. Robins. 2007. "The identification of synergism in the sufficient-component-cause framework." Epidemiology 18:329-339.

Winship, Christopher and Michael Sobel. 2004. "Causal Inference in Sociological Studies." in Handbook of Data Analysis, edited by M. Hardy and A. Bryman: Sage Publications.

Wright, Erik Olin, Andrew Levine, and Elliott Sober. 1992. Reconstructing Marxism: Essays on Explanation and the Theory of History. London: Verso.