THE RED HERRING OF PROBABILITY RAISING

NAFTALI WEINBERGER

Abstract. Theorists of probabilistic causality viewed causation as probability raising relative to particular contexts. In contrast, more recent graphical theories do not specify whether a cause raises or lowers the probability of its effect as part of the causal representation, but enable one to infer such quantitative facts from the joint probability distribution and additional causal assumptions. This difference between the accounts may seem minor, but here I argue that the emphasis among probabilistic theorists on probability raising reflected their not having an adequate theory of confounding and thus of the relationship between causal and probabilistic claims. The graphical account of confounding clarifies why causal relationships need not be identified with particular probabilistic relationships in particular populations, and thus why many of the earlier debates about probability raising in retrospect no longer appear to reflect substantive philosophical differences.

1. Introduction

Whatever happened to "positive causal relevance"? Back in the heyday of probabilistic theories of causation, one of the key tasks for an account was to give the conditions under which a cause is positively relevant to its effect. Broadly speaking, this was explicated in terms of probability raising across a set of relevant context, though there were disputes regarding how to specify the contexts, and regarding the range of contexts across which a cause must raise the probability of its effect to count as a positive causal factor. Nowadays, there are few people who claim to be providing a probabilistic theory of causation. It is more common to present causal theories relying on graphical models (Pearl, 2009; Spirtes et al., 2000), which, depending on how you look at it, have either rendered probabilistic theories obsolete, or count as the successful culmination of the probabilistic project. Like the earlier accounts, graphical approaches to causation trace the systematic relationships between causal and probabilistic concepts. Yet many projects from the earlier literature – such as that of defining positive causal relevance – no longer seem central to providing an account of causation. So the question is: what changed? Why is it that certain debates that at one point seemed crucial to understanding causation no longer seem central, when they are discussed at all?

Here I argue that the reason that theorists of probabilistic causality were so preoccupied with defining positive causal relevance was that they lacked an adequate account of confounding. This proposal will initially sound implausible, since probabilistic theorists had most of the elements of a proper account of confounding. They were generally aware of

Date: January 9, 2025.

which types of variables one needs to condition on to eliminate confounding, and were also aware – though one finds some confusion on this point – that one need not hold fixed *all* background factors in order to eliminate confounding. Nevertheless, without the general account of confounding provided by graphical causal models, they missed something essential. On the graphical account, eliminating confounding amounts to using one's causal assumptions to determine whether a particular probabilistic expression provides an unbiased estimate of a causal effect. Crucially, one need not make any assumptions about this probabilistic expression – and in particular about whether the causal variable is positively or negatively correlated with the effect variable – in order to establish whether it provides such an unbiased measurement. In contrast, the earlier theorists, as a result of linking causation to positive causal relevance, viewed confounding as a matter of finding a privileged causal relationship (i.e. probability raising) that corresponds to the causal effect. While probabilistic theorists gradually relaxed the requirement that causes be probability raisers, they never abandoned the assumption that an account of causation should explicate causal claims in terms of particular probabilistic relationships that obtain in particular contexts and, correspondingly, that the account must pick out the relevant contexts. This explains both why they saw it as crucial to give an account of how facts about probability raising vary across contexts, and also why it no longer is necessary to do so in order to give an account of causation.

That graphical accounts are more technically advanced than their probabilistic predecessors is not by itself surprising. Yet there has been little discussion regarding whether this technical progress has led to a different philosophical understanding of the relationship between causation and probability. My argument here aims to show that it has and to identify the technical developments that led to this improvement. Doing so is important for seeing the philosophical contributions of graphical methods. My discussion also calls for reinterpreting key debates from the literature on probabilistic causality. A consequence of my discussion is that these debates would have been avoided given an adequate account of confounding.

The paper is organized as follows. Section 2 provides a brief historical account of the probabilistic causality project, and explains why the discussion focuses on non-reductive probabilistic accounts. Section 3 highlights the various debates among probabilistic theorists in which probability raising played a central role. Section 4 explains why these debates no longer are pressing in light of graphical models. Section 5 argues that the earlier emphasis on probability raising reflected and helped perpetuate an inaccurate understanding of confounding, and thus of the relationship between causal and probabilistic claims. Section 6 concludes.

2. Historical Context

Early theories of probabilistic causality were reductive, in that they sought to wholly explicate causal claims using non-causal concepts. Cartwright (1979) argued that no reductive account is available, but that one can nevertheless give an informative account of the relationship between causal and probabilistic claims. Almost all probabilistic accounts

after Cartwright are non-reductive, and graphical accounts resemble the later accounts in this respect. Here I am interested in the relationship between graphical approaches to causality and *non-reductive* probabilistic accounts. In focusing on these accounts, I do not presuppose that there cannot be a reductive account. Whether there can be is not as settled as is commonly presumed, and Spohn (2012) offers a reductive probabilistic account meriting further investigation. If a reductive account should succeed, then the philosophical disagreement between such an account and non-reductive ones will be evident enough. Whether non-reductive probabilistic and graphical theories differ in their philosophical commitments is less obvious, and is the topic of the present paper.

To illustrate the difference between reductive and non-reductive accounts, I begin with Patrick Suppes' non-reductive account (Suppes, 1973). On his account, event C is a *prima facie* cause of E iff (i) C occurs before E and (ii) C raises the probability of E (over ¬C, or, equivalently, over the unconditional probability of E). It is also assumed that C's probability is greater than 0. Not all prima facie causes are genuine causes. My drinking a strong blond Belgian beer now makes it more probable that I will be happy now and also more probable that tomorrow I will have a headache, but my being happy is not the cause of my headache.[1] Here the correlation between my being happy and of my having a headache is explained by my drinking the beer, and if this is the only (and most proximate) common cause of these effects, the latter will be probabilistically independent conditional on my drinking the beer. Metaphorically, my drinking the beer *screens off* the probabilistic relationship between its effects. The crux of Suppes' account is that a prima facie causal relationship between C and E is a genuine causal relationship iff there is no factor F prior to C that screens off C from E.

As Suppes makes clear, to give an account of probabilistic causality, one cannot explicate causal relations simply in terms of C preceding E and raising its probability. One needs to further specify the *contexts* relative to which C's raising the probability of E indicates that C causes E. The context must include variables that confound the causal relationship between C and E, such as common causes. To give a reductive account, it must be possible to specify which variables must (and must not) be included in the context without appealing to causal facts. For instance, one cannot say that all common causes of C and E must be held fixed, but might instead require that all variables temporally prior to C be held fixed. Non-reductive accounts, in contrast, help themselves to assumptions regarding the causal relationships between certain variables in determining which variables must be included in a context. For example, Eells (1991) requires including all causes of E that are not effects of C. Non-reductive accounts do not allow one to get any causal knowledge without prior causal knowledge, but are nevertheless helpful for combining one's probabilistic knowledge with one's prior causal knowledge to make novel causal inferences.

[1] I am not here concerned with claims about actual causation, which entail that the cause and effect occur. This statement should be understood in terms of the relationship between the variables of drinking a beer (or not), being happy (or not), and having a headache (or not). While probabilistic theories take events as the causal relata, in graphical accounts the relata are *variables* (denoted by italicized uppercase letters). Events may be represented using dichotomous variables for whether they occur, where C and ¬C are denoted by (e.g.) $C = 1$ and $C = 0$, respectively.

Many of the debates in the later literature on probabilistic causality concern which variables need to be included in the background context in evaluating the effect of some C on some E. Consider the effect of being engaged in military combat on later experiencing post traumatic stress disorder (PTSD). The amount by which combat will increase the probability of PTSD varies greatly across individuals, based on background factors such as prior history of trauma, military unit, post-deployment social relationships, and an indefinite number of other background factors. Even if one can eliminate confounding, it will still be the case that combat influences PTSD differently across different individuals, and could potentially even lower its probability for a subset of individuals (by making them better prepared to deal with future stress). Theorists of probabilistic causality devoted significant energy to the question of which background contexts need to be considered. For instance, should we say that combat experience causes PTSD even if it does not raise its probability across *all* background contexts? And must the background context include all background factors that make a difference to the causal relationship? If one omits such factors, then one may still be able to eliminate confounding, but one will at best be able to describe an average effect (since the probabilistic relationship between C and E will be an average across the different values of the background variables that are not held fixed).

Claims about whether a cause raises or lowers the probability of an effect concern *causal role*. Such claims are more specific than those specifying whether a cause is relevant to its effect without being more precise about the relationship. Theorists in the probabilistic literature saw their task as that of explicating claims regarding causal role. For instance, in Ellery Eells' (1991) account, cause C is *positively relevant* to its effect E in some population P if C raises the probability of E across all contexts in the population. Whether this was the correct definition was a point of contention (see e.g. Skyrms, 1980). More generally, debates in the literature regarding background contexts all reflect a preoccupation with causal role. It is because probabilistic theorists were concerned with explicating role and causal roles vary across background contexts that so much attention is given to background contexts.

In a widely cited paper, Hesslow (1976) claims:

> The basic idea in Suppes' theory is of course that a cause raises the probability of its effect, and it is difficult to see how the theory could be modified without upholding this thesis. (291)

This illustrates that the probability raising requirement was seen as indispensible to theories of probabilistic causation. Contrast this with Hitchcock's claim in his encyclopedia article on probabilistic causality, that "causal relevance [rather than more specific facts about causal role] is really the basic metaphysical concept" (Hitchcock, 2018). As we will see, Hitchcock's claim is supported by the methodology of graphical causal inference. What remains to be determined is why probabilistic theorists saw probability raising as indispensable to a causal account, and which developments enable graphical accounts to evade thorny debates about causal role.

Before proceeding, it is important to acknowledge that theorists of probabilistic causality were simultaneously engaged in several projects, some of which will not be relevant to

what follows. In addition to spelling out the relationships between causation and probability, they were also involved in the metaphysical project of giving an account of irreducibly probabilistic causal relationships and in the linguistic project of understanding the truth conditions of causal generalizations . Regarding the metaphysical project, Hausman (2010) has argued that these theorists mistakenly assumed that a probabilistic theory of causal generalizations should rely upon or be relevant to understanding causality in indeterministic contexts. The following will illustrate why there can be probabilities in causal models even if the world is deterministic. I will also have little to say about the linguistic project, although it is important to acknowledge that probabilistic theorists focused on probability raising in part because they were explicating claims of the form "C causes E", which entail that the cause promotes its effect. While concern with linguistic analysis may explain why they initially explicated causation in terms of probability raising, it cannot fully account for the extent to which classifications of causal role were emphasized and debated. The next section will catalogue the extent to which probability raising was at the core of many debates in this literature, none of which directly concerns linguistic usage.

3. Debates Over Probability Raising

In this section I illustrate the extent to which probabilistic theorists were preoccupied with causal roles and with how they vary across contexts. Concerns about causal role are central to a host of core debates in this literature, including those concerning contextual unanimity, interaction effects, average effects, and multi-path effects.

Cartwright (1979) provides an influential account of which variables must be included in the background context. She specifies the conditioning set $\mathbf{K}$ relative to which C's raising the probability of E indicates that C causes E, as follows:

(1) $\mathbf{K}$ must only include only causes of E or of $\neg E$
(2) C itself it not in $\mathbf{K}$
(3) $\mathbf{K}$ includes all causes of E (and $\neg E$) other than C, with the exception of:
(4) Causes of E that are effects of C, which are excluded from the conditioning set

According to Cartwright, C causes E iff C raises the probability of E relative to all values of all the variables in $\mathbf{K}$. It is natural to see Cartwright as providing conditions for which variables must be held fixed in order to eliminate confounding, though note that condition (3) requires holding all causes of E fixed, and this is not necessary to eliminate confounding.

For what follows, it will help to highlight two distinct ways in which Cartwright's definition requires that one hold all background factors fixed in establishing causal claims. First, condition (3) requires one to include all causally relevant factors in the context. This condition ensures that causal effects are established relative to homogenous background contexts, and that any case in which a cause does not determine the value of its effect reflects genuine indeterminism. Additionally, Cartwright requires that C raise the probability of E across *all* background contexts. This is a type of contextual unanimity requirement. Skyrms (1980) offers a weaker version on which causes must raise the probabilities of their effects in some contexts and lower them in none.

The obvious issue with requiring contextual unanimity is that most causal relationships obtain only relative to some background contexts. Causal relationships typically do not obtain in a vacuum but require certain other factors to obtain. Paradigmatically, striking a match causes it to light, but only in the presence of oxygen. When the magnitude of the effect of X on Y depends of the value of a third variable Z, we say that X and Z *interact* in their effect on Y. Contextual unanimity, in its strongest form requiring that a cause must raise the probability of its effect across all contexts, implies that striking matches does not cause them to light. Cartwright deals with interaction by saying that there are different causal laws for each combination of causes. That is, there are different causal laws for the results of striking a match in the presence of oxygen and for striking it in the absence of oxygen, and so on. So even if we are forced to say that striking a match does not cause it to light, we may be able to say that striking a match in the presence of oxygen does (provided that these causes also raise the probability of lighting conditional on additional factors).

Eells (1986) provides an alternative way to model interaction. Eells relativizes claims about causal role to populations. In the match case, Eells would advocate distinguishing between the populations in which there is and there is not oxygen. Striking the match may raise the probability of its lighting in all contexts for the population with oxygen even if it does not do so across all contexts in the mixed population of strikings in the presence and absence of oxygen.

This marginal dispute over modeling interaction provides an opportunity to say more about Eells' talk of populations. Eells (1991) relativizes his account to token populations that are characterized as being of a particular type. The relativization to populations can be understood in terms of the necessity of characterizing which concrete entities are represented using a given probability distribution. One can refer to a population without specifying the distribution of the background factors characterizing it. An example of a population might be that of New Yorkers between 20 and 40. The relativization of causal claims to populations makes sense insofar that combining two arbitrary samples may not yield a sample from which one can straightforwardly infer conditional probabilities from relative frequencies. Yet in Eells' treatment of interaction, his characterization of populations as types gives him an adjustable parameter for achieving contextual unanimity. That is, whenever a cause raises the probability of its effect in a population only for those individuals with some factor X, one can get unanimity by dividing the population into two subpopulations based on which individuals have X. Naturally, it is often helpful to divide up populations in this way, especially where X is a salient causal factor. But as a strategy for preserving contextual unanimity, it is only valuable if one is already committed to causes having an unambiguous role within a population.

Whether it is in fact important that causes have such a role is the source of a debate between Eells and John Dupré. Dupré (1984) argues that causes need not be evaluated relative to a homogenous background context and thus that the context need not include all causally relevant factors. Absent such homogeneity, the variation in non-fixed factors may lead to variation in the effect of C on E within a population, and the most one can measure is an average effect. Eells' dismisses the average effect as "a sorry excuse for a

causal concept" (Eells, 1987, p. 113). A common explanation for Eells' pity is that he is trying to provide an account of causal laws, and laws should not depend on the contingent frequencies of background factors (Hitchcock, 2003, p.15). We will return to this debate later, since Dupré is plausibly understood as rejecting the importance of causal role.

A further domain in which the emphasis on positive causal relevance led to extended debate is in discussions of causes that influence their effects in multiple ways (Hesslow, 1976; Dupré, 1984; Otte, 1985; Cartwright, 1989)). In Hesslow's (1976) paradigmatic example, taking birth control pills reduces one's chance of thrombosis by reducing one's chance of getting pregnant, which is a risk factor for thrombosis. But it also increases one's chance of getting thrombosis via producing some blood chemical. Here I will not go into the particulars of these debates, which are covered by Weinberger (2017). The salient point for our purposes is that they concerned whether birth control has an unambiguous positive or negative effect of thrombosis, and whether we can account for the way that birth control has positive influence via some avenues, and negative influence via others.

A final way in which the emphasis on causal role figured probabilistic accounts was in motivating the distinction between type and token causation (Sober, 1984). A key motivation for the distinction was Rosen's example involving a squirrel kicking a golf ball into a hole in a way that lowers the probability of its going in the hole, but which results in the ball going in nonetheless (Rosen, 1978). The type/token distinction aims to resolve the problem by saying that while the kick is a type-level negative factor of the ball's going into the whole, it is nevertheless a token positive cause of it. The type/token labels have been in fact used to refer to multiple different distinctions related to causal heterogeneity/homogeneity, deterministic/irreducibly stochastic causal explanation, and causal tendencies/actual causes. Here I'll say little about these distinctions beyond noting that I am concerned here with facts about causal tendencies, whether in individuals or populations, and make no assumptions about whether C and E are instantiated.

In this section, we have seen that the emphasis among theorists of probabilistic causation on probability raising shaped the core debates that occupied them in the eighties and nineties. What remains unclear is why they saw probability raising as so crucial. In the next section, I will explain why probability raising is not central in graphical accounts of the causal relationships between variables, and then use the difference between the approaches to shed light on what hindered probabilistic theorists from generalizing their accounts to deal with causal relevance.

4. Generalizing Beyond Probability Raising Using Causal Graphs

Graphical causal models represent causal hypotheses using directed graphs in which the nodes are random variables and the directed edges (i.e. arrows) indicate direct causal relationships. For the moment we will leave the concept of direct causation as an undefined primitive. As is common, I will focus on directed *acyclic* graphs (DAGs) in which one cannot get from any node back to itself via a series of connected arrows all pointing in the same direction. DAGs have revolutionized causal inference through the development

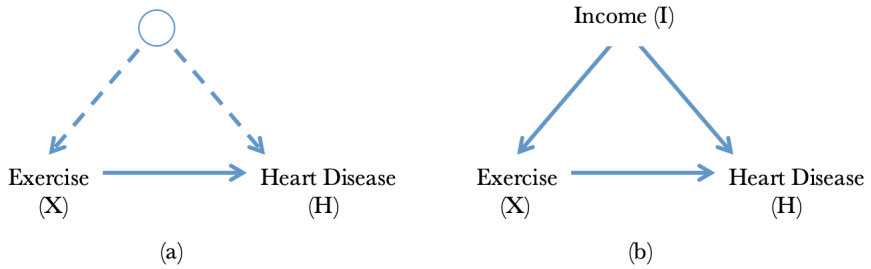

FIGURE 1

of algorithms by which one can choose among causal hypotheses given a probability distribution (Spirtes et al., 2000). These search methods have received considerable attention from philosophers, and much of it has focused on the appropriate bridge principles for relating causal hypotheses to probabilistic facts about conditional independence (Hausman and Woodward, 1999; Cartwright, 2002; Andersen, 2013; Weinberger, 2018). But, perhaps surprisingly, these will not be emphasized here. Rather, I will focus on the notion of causal *identifiability* (Pearl, 2009, definition 3.2.4). Identifiability concerns whether, given knowledge of the correct DAG, a particular causal quantity can be uniquely determined from the probability distribution. The notion of identifiability is crucial to providing an adequate account of confounding.

To illustrate identifiability, consider the simple DAG in figure 1(a) in which exercise (X) influences heart disease risk (H), but where there is also an unmeasured common cause of these variables (represented by a dashed bidirected arc – DAGs with bidirected arcs are called *semi-Markovian models*). Here the effect of X on H is *not* identifiable, since one cannot determine whether any correlation between X and H is the result of the common cause or the direct causal relationship. In contrast, suppose (for simplicity) that these variables have a single common cause, income (I)(Figure 1b). Now the effect of X on H is identified by the distribution $P(H|X, I)$. The correlation between X and H for a given value of I corresponds to the effect of X on H for that value of I.

When a probabilistic expression identifies the effect of C on E, one can use it to infer how the probability of E would vary given *ideal interventions* on C. Roughly, ideal interventions on C determine C's value in such a way that C no longer depends on its direct causes (other than the intervention) and any influence of the intervention on C's effects are only via C. $P(E|do(C))$ denotes the probability of E given and intervention on C. Whenever some probabilistic expression $P(Y|X, \mathbf{V})$ identifies the effect of X on Y, $P(Y|X, \mathbf{V}) = P(Y|do(X), \mathbf{V})$. It is increasingly common to explicate causal relationships between variables in terms of the notion of an ideal intervention (Woodward, 2003). That is, C causes E if it is possible to change the value (or distribution) of E via an intervention on C. Everything I say here is compatible with an interventionist account of causation. My reason for not emphasizing interventions in my discussion is to avoid giving the impression

that they serve as a key difference between graphical and probabilistic approaches. When an effect is identifiable from the probability distribution and the DAG, one can measure it without intervening. This is not to say that one can infer causal relationships without substantive causal knowledge, but rather to note that later theorists of probabilistic causality were willing to help themselves to such knowledge.

Let's now see how DAGs allow for a general account of identifiability. We've noted that the effect of C on E is not identified if one does not condition on a common cause and can be identified conditional on their common cause. Here it helps to distinguish between *causal paths* consisting of a set of connected arrows all going in the same direction from C to E, and *non-causal paths* in which not all of the connected arrows go in the same direction. One way to understand why the effect of X on H is not identified in figure 1(a) is that the non-causal path from X to H via the unmeasured common cause "transmits" probabilistic information via a path other than the causal one. This is why it is underdetermined whether the probabilistic relationship between X and H is due to the causal or non-causal path. Conditioning on the common cause can be seen as blocking the transmission of probabilistic influence via this path.

While paths with common causes that one does not condition on transmit probabilistic influence, paths with common effects will *not* transmit probabilistic influence *unless* one conditions on them. Suppose that one cause of restaurant popularity is having good food and that another independent cause is having a good view. Conditional on a restaurant's being popular, there will be a negative correlation between its having a good view and its having good food. The reason for this, roughly, is that since there are two factors that could explain why a restaurant is popular, learning that one factor obtains and potentially explains the popularity lowers the probability that it is due to the other factor. As a result, if it were the case that, somewhat bizarrely, a restaurant's having a good view had some effect on the quality of its food (say, via inspiring the chefs), $P(FoodQuality|View, Popularity)$ would not identify the effect of the view on food quality. To identify this effect, one would have to *not* condition on *popularity* (that is, to consider a sample with both popular and unpopular restaurants).

These examples provide an informal sense of how one can determine whether a causal quantity is identifiable by considering whether there is a non-causal path that transmits probabilistic influence between the cause and the effect. We can make this more precise using the concept of *d-separation* (Pearl, 2009):

d-separation: A path [set of connected arrows] is *d-separated* variable set $\mathbf{Z}$ just in case:

(a) The path contains a triple $i \rightarrow m \rightarrow j$ or $i \leftarrow m \rightarrow j$ such that m is in $\mathbf{Z}$, or

(b) The path contains a collider $i \rightarrow m \leftarrow j$ such that m is not in $\mathbf{Z}$ and no descendant of m is in $\mathbf{Z}$. [n is a *descendant* of m if there is a path from m to n consisting of directed arrows all going in the same direction])

d-separation is a more precise characterization of what it means for a path to *fail* to transmit probabilistic influence. d-separation is defined here as a property of paths. Two variables are d-separated (simpliciter) by $\mathbf{Z}$ if and only if they are d-separated by $\mathbf{Z}$ along all paths. Two variables are *d-connected* by $\mathbf{Z}$ if and only if they are not d-separated.

A sufficient (but not necessary) condition for identifying the effect of C on E conditional on variable set $\mathbf{X}$ is that $\mathbf{X}$ d-separates C from E along all non-causal paths. This is why in the cases considered the effect was identifiable as long as one conditioned on any common causes, but no common effects. There is no necessary and sufficient condition for picking out the variables that one must condition on in order to identify a causal relationship, but Pearl (1995) provides a sound and complete procedure for determining when a quantity is identified given a DAG and joint probability distribution over its variables.

The notion of identifiability enables one to give an analysis of causation that is in the spirit of the earlier non-reductive probabilistic analyses. X is causally relevant to Y if and only if X is probabilistically relevant Y relative to at least one background, $V = \{v_1, ..., v_n\}$, which includes variables blocking all and only the non-causal paths between X and Y. Given that in introducing identifiability I assumed knowledge of the causal relationships, how does the notion of identifiability account for the difference between $X \rightarrow Y$ and $Y \rightarrow X$? The answer is that changing the direction of the arrow can change the variables one needs to condition upon in order to identify the causal relationship. (Cartwright, 1979, p. 429) already appreciated the significance of the fact that $X \rightarrow Y$ and $Y \rightarrow X$ are evaluated relative to different background factors, although she did not talk in terms of identifiability.

The concept of identifiability allows for a clear-cut distinction between the causal and statistical aspects of causal inference. Whether a quantity is identifiable depends solely on the qualitative causal assumptions about the paths by which the variables in the relationship are linked. Identifiability entails that the quantity can in principle be estimated using statistical methods by which one infers the probability distribution from a finite sample. The causal model places no constraints on the functional relationship by which an effect depends on its causes. It is silent for example, regarding whether the causes of an effect interact or contribute additively. When a quantity is identified by a probabilistic expression, it is possible to get an *unbiased effect measurement* of it using statistical methods. While it is always possible that one's sample will be unrepresentative of a broader population, identifiability ensures that there are no factors that systematically distort one's estimate of the desired quantity. Factors that bias an effect measurement are confounders.

Average effects can be identified even when one does not model variables that influence the magnitude of the effect. Consider again the effect of exercise on heart disease risk and suppose (for the purposes of illustration) that there are no confounding variables. Within a population, the effect of exercise on heart disease could vary among individuals with different family histories (among many other risk factors). Suppose that exercising more than 30 minutes a day reduces the risk of heart disease by .2 for individuals with a family history of heart disease, and by .1 for individuals without such a history. The population contains some unknown mix of individuals with and without such a family history. So how can we measure the average effect in the population? As long as our sampling process is random with respect to family history, our sample will contain the same percentage of individuals with and without a history as in the general population, and the effect in a population is just a weighted average over the effects in the individuals. For instance, if .25

of the population has a family history, then the average effect will be (.25)(.2)+(.75)(.1) = .125. And so on for other background factors. Provided one eliminates confounding, average effects pose no problem for causal identifiability. This is not to deny, of course, that it is often useful to include more factors in order to measure subpopulation-specific effects.

Eliminating confounding amounts to there being a probabilistic expression that, given the probability distribution and the DAG, identifies the causal effect. The graphical account of identifiability reveals that whether confounding can be eliminated depends solely on the qualitative assumptions embedded in the graph. Moreover, in representing a causal relationship in the graph, one need not distinguish between positive and negative effects, or to worry about average effects or interaction. As long as an effect is identifiable, the precise relationship between an effect and its causes can be inferred from the probability distribution, and the identifying probabilistic relationships provide an unbiased measurement of the effect whether or not it is homogenous across or within populations.

In giving a causal model, one can be almost entirely silent regarding the background factors for the causal relationships in the model. The influence of the background factors on the causal effect is already reflected in the probabilistic expression identifying the effect. To the extent one needs to make any assumptions about background factors, one needs to assume that there is *some* set of background factors relative to which the cause is probabilistically relevant to the effect. But these need not be background factors that obtain in the population being considered. If one assumes that some C influences some E, but considers a population in which C influences E via cancelling paths, or in which the average effect across subpopulations is zero, then one will correctly identify C as having zero net average effect on E. Such cases pose potential epistemological problems for causal search (Andersen, 2013), since we may not discover the relevant causal relationship. But bracketing the issue of how we discover that C causes E, the claim $C \rightarrow E$ comes with no commitments about the probabilistic relationship between C and E any particular background context.

5. Why Did Probabilistic Theorists Emphasize Probability Raising

Given an account of causal identifiability, there is no need to give a classification of causal roles in order to explicate causation. One can determine whether a causal relationship is identifiable without making any assumptions about causal role, and when it is identifiable the causal role (if any)[2] may be read off of the probability distribution. My proposal is that that earlier theorists were so preoccupied with issues of causal role and effect heterogeneity because they lacked a general account of identifiability, and thus of

[2]For non-dichotomous variables, the causal role may be indeterminate. For example, if some increases in a cause increase the probability of the effect, while further increases lower it, the cause is neither positive, negative, nor neutral. Hitchcock (1996) has already explored issues in generalizing probabilistic theories to non-dichotomous variables.

confounding. To defend this proposal, I need to argue that probabilistic theorists' account of confounding was not merely less general, but in fact led them to misconstrue the relationship between causation and probability.

To this day, one can find the concept of "confounding" used in two different ways. On a widespread understanding of confounding, confounders are causal explanations of an effect other than that proposed by an experimenter. In this sense, to criticize a study of the effect of a college degree on future income on the basis that it does not control for the confounder of family wealth is to suggest that family wealth would provide an alternative explanation for why college educated individuals have higher incomes. This way of talking suggests that there is a particular type of effect that we are trying to uncover – in this case a positive one – and that confounders are alternative factors that might raise the probability of the effect. We can refer to this understanding of confounding as the *colloquial* one. This contrasts with the graphical understanding on which confounders are variables that bias one's effect measurement – whatever the quantitative effect might be.

While the contrast between the colloquial and graphical notions is subtle, adopting the former leads to misconceptions about the relationship between causation and probabilities. Linking causal effects to particular probabilistic relationships amounts to requiring that causal hypotheses be specified relative to particular sets of background factors. This in turn leads one to worry about heterogeneity in background factors, whether across or within populations, and to see alternate or interactive causes of an effect as obscuring an effect's "true" causal role. I propose that the focus on such issues among probabilistic theorists results from their having a colloquial rather than a graphical understanding of confounding, and that this can be traced back to the original gloss of causation as positive causal relevance. For causation to be probability raising, causal claims must be specified relative to particular sets of background factors. While probabilistic theorists were willing to relax the assumption that causes must always raise the probability of their effects, they never abandoned the idea that explications of causal claims must specify facts about causal role, and thus must be relative to some set of background factors.

Although I have thus far emphasized that probabilistic theorists emphasized probability raising *because* they lack an account of confounding, the relationship goes both ways. That is, I am claiming both that given an adequate account of confounding probabilistic theorists would not have been so obsessed with giving an account of causal role, and also that their identification of causal relationships with particular types of probabilistic dependence (such as probability raising) hindered them from developing an adequate account of confounding.

While probabilistic theorists were aware that confounding and heterogeneity were distinct issues, these were regularly conflated. Here, for example, is a quotation from Eells' response to Dupré:

> Because of the well-known phenomenon of "spurious correlation," it is not enough that C simply raise the probability of E on average. By insisting

> that the cause raise the probability of its effect in *every* "background context," we control for correlated factors that may be responsible for spurious correlations. (Eells, 1986; p. 105, emphasis in original).

Eells treats contextual unanimity as a solution to the problem spurious correlation – otherwise known as confounding. Similarly, one of Cartwright's (1979) initial examples of confounding encourages the conflation of confounding and heterogeneity. In it, the effect of uranium on a Geiger counter is confounded in that whenever the uranium is absent, radioactive polonium is present. Following Cartwright's advice to evaluate the effect of uranium (vs. no uranium) relative to contexts in which plutonium is held constant will yield an unbiased effect measurement. Yet it is a feature of the case that uranium raises the probability of activating the Geiger counter both in contexts with and without the plutonium (and Cartwright's account requires one to consider both). This encourages one to think of causes as having a characteristic effect across contexts that may be obscured by confounders.[3] This tendency is only further encouraged by her later work on causal capacities (Cartwright, 1989).

Intriguingly, Cartwright does at one point acknowledge that probability raising is not essential to her account. She notes (1979, p. 426) that in Salmon's account, the explanatory relationship is understood as probability changing rather than probability raising. While she in principle remains open to alternative probabilistic relationships, what matters to her is that whatever relationship one chooses, whether it obtains should not be a matter of the way the situation is described. The problem is that probabilities are notoriously sensitive to the set of variables modeled – this is known at the *reference class problem*. Cartwright avoids making probabilistic causal relationships from being reference-class dependent by relying on causal facts in specifying the reference class. Specifically, in describing the situation in which the relevant statistical relationship obtains, we must include "all and only the other causally relevant features" (426).

The graphical account of confounding reveals why variation in background contexts should not be characterized as a reference class problem. The causal relationship between some C and E will *genuinely* vary across contexts, and – assuming that one has conditioned on a variable set that is sufficient to eliminate confounding – one will identify the correct causal relationship for the distribution of background factors in the population. The reason that our description of the relationship between C and E varies across contexts is because the relationship itself varies. This variation does not correspond to some potentially problematic dependence of causal facts on the way we describe the population. But thinking of some particular relationship such as probability raising as capturing the genuine causal relationship makes it seem like there is a reference class problem, since one's preferred relationship will only obtain relative to certain background factors. Accordingly, one will see it as essential to understanding causation to specify what those factors are.

[3]Cartwright also notes that while uranium lowers the probability that the counter will click, "when the uranium has been drawn and the Geiger counter does register a large number of clicks, it is the uranium that cause" (422). This invites a further conflation between establishing causal relevance and providing post-hoc attributions of responsibility for an instantiated effect.

The idea that eliminating confounding is a matter of isolating a set of contexts in which an invariant probabilistic expression obtains would explain why they were so preoccupied with categorizing causal roles and why interaction and effect heterogeneity would seem problematic. Moreover, it is hard to see how to avoid identifying causal relationships with some set of background factors without a *general* account of confounding. Consider cases of *Simpson's paradox*, in which there is a positive (negative) correlation in a population, but no positive (negative) correlations in the subpopulations derived by partitioning the population based on additional variables. Such cases worried probabilistic theorists, since if a correlation entirely disappears when partitioning, this suggests that it does not correspond to a causal effect. Given the graphical account of confounding, it is trivial to show that one cannot have a positive cause of an effect that is not a positive cause in at least some subpopulations (Weinberger, 2015). Accordingly, if we correctly identify an effect, it will not disappear upon partitioning. Without a general account of identification, however, one can have no such assurances. That is, one cannot, in general, determine when conditioning on a set of variables is sufficient for eliminating confounding and thus be assured that one has an effect that will not be revealed to be spurious on further partitioning. Without an assurance that one's measured effects will be preserved under further partitioning, the only way to guarantee that one's effects are genuine is to maximally partition the population based on all causes of the effect.

The idea that causal relationships correspond to invariant causal roles accounts for why Eells saw the average effects as "a sorry excuse for a causal concept". If causal concepts must be specified relative to a set of background factors, average effects will be faulted to the extent that they are defined for a population without characterizing the way that the relationship varies across background factors. Where Eells goes wrong is in assuming that causal effects must be specified relative to a set of background factors. But a proper account of confounding is necessary in order to demonstrate that they need not be.

My analysis here deviates from the dominant contemporary reading of the debate between Dupré and Eells, which is offered by Hitchcock (2003). He suggests while Dupré was interested in the average effect because of its interest for policymakers, Eells cared about effects in homogenous populations because he is giving an account of causal laws. An initial point to make here is that Dupré does not himself provide an account of confounding, and Eells' justifiably dismisses his unhelpful proposal that effects should be evaluated relative to a "fair sample". More importantly, we need to ask why it would be valuable for Eells to distinguish causal laws from supposedly non-lawful causal generalizations. Given the graphical account of confounding, it is clear that facts about causal role can vary across contexts and that this is unproblematic. Eells' attempt to gerrymander populations into ones with homogenous effects does not eliminate this variation, but tries to define it out of existence.

Of course, there is a philosophical tradition of linking explanation to laws, understood as fully invariant generalizations. But as (Woodward, 2003, Ch. 6) has emphasized, in adopting a causal theory of explanation we should switch to thinking of invariance as coming in degrees. Additionally, there are two types of invariance: the invariance of the probabilistic relationship between a cause and effect to interventions on the cause and the

invariance of this relationship to changes in background factors. Probabilistic theorists emphasized the second type of invariance. The upshot of our discussion of identifiability is that in giving an account of the relationship between causes and probabilities, it is the first type of invariance that is fundamental. We should reject any proposal treating causal relationships that generalize across a wide set of background factors as any more legitimately causal than those that do not.

6. Conclusion

Graphical account of causation are not merely a technical advance over probabilistic ones, but a conceptual advance as well. As a result of not having a general theory of confounding and identifiability, theorists of probabilistic causality misconstrued the relationship between causation and probability. Since contemporary graphical models enable one to systematically relate causes to probabilities in a manner that builds on work done in developing probabilistic accounts, it is tempting to treat them as merely providing more sophisticated probabilistic accounts. But they ultimately undermine the idea that causation should be understood in terms of a probabilistic relationship that obtains in particular contexts. The causal content of graphical models is given not by the particular probabilistic relationship that identifies an effect, but by the qualitative relationships that determine whether that effect can be identified. Although probabilistic dependence relationships will in certain circumstances be causal relationships as well, probability theory is the wrong place to look in giving an account of what makes these relationships causal.

References

H. Andersen. When to expect violations of causal faithfulness and why it matters. *Philosophy of Science*, 80(5):672–683, 2013.

N. Cartwright. Causal laws and effective strategies. *Noûs*, pages 419–437, 1979.

N. Cartwright. *Nature's Capacities and their Measurement*. Clarendon Press, 1989.

N. Cartwright. Against modularity, the causal markov condition, and any link between the two: Comments on hausman and woodward. *The British journal for the philosophy of science*, 53(3):411–453, 2002.

J. Dupré. Probabilistic causality emancipated. *Midwest Studies in Philosophy*, 9(1):169–175, 1984.

E. Eells. Probabilistic causal interaction. *Philosophy of Science*, 53(1):52–64, 1986.

E. Eells. Probabilistic causality: reply to john dupré. *Philosophy of Science*, 54(1):105–114, 1987.

E. Eells. *Probabilistic causality*, volume 1. Cambridge University Press, 1991.

D. M. Hausman. Probabilistic causality and causal generalizations. In *The place of probability in science*, pages 47–63. Springer, 2010.

D. M. Hausman and J. Woodward. Independence, invariance and the causal markov condition. *The British journal for the philosophy of science*, 50(4):521–583, 1999.

G. Hesslow. Two notes on the probabilistic approach to causality. *Philosophy of science*, 43(2):290–292, 1976.

C. Hitchcock. Of humean bondage. *The British Journal for the Philosophy of Science*, 54 (1):1–25, 2003.

C. Hitchcock. Probabilistic causation. *Stanford Encylopedia of Philosophy*, 2018.

C. R. Hitchcock. Farewell to binary causation. *Canadian Journal of Philosophy*, 26(2): 267–282, 1996.

R. Otte. Probabilistic causality and simpson's paradox. *Philosophy of Science*, 52(1): 110–125, 1985.

J. Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995.

J. Pearl. *Causality*. Cambridge university press, 2009.

D. A. Rosen. In defense of a probabilistic theory of causality. *Philosophy of Science*, 45 (4):604–613, 1978.

B. Skyrms. Causal necessity: a pragmatic investigation of the necessity of laws. 1980.

E. Sober. Two concepts of cause. In *PSA: Proceedings of the biennial meeting of the philosophy of science association*, volume 1984, pages 405–424. Philosophy of Science Association, 1984.

P. Spirtes, C. N. Glymour, R. Scheines, D. Heckerman, C. Meek, G. Cooper, and T. Richardson. *Causation, prediction, and search*. MIT press, 2000.

W. Spohn. *The laws of belief: Ranking theory and its philosophical applications*. Oxford University Press, 2012.

P. Suppes. A probabilistic theory of causality. 1973.

N. Weinberger. If intelligence is a cause, it is a within-subjects cause. *Theory & Psychology*, 25(3):346–361, 2015.

N. Weinberger. Path-specific effects. *The British Journal for the Philosophy of Science*, 70 (1):53–76, 2017.

N. Weinberger. Faithfulness, coordination and causal coincidences. *Erkenntnis*, 83(2): 113–133, 2018.

J. Woodward. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press, 2003.