Foundations for a Probabilistic Theory of Causal Strength*

Jan Sprenger†

February 3, 2016

Abstract

This paper develops axiomatic foundations for a probabilistic-interventionist theory of causal strength. Transferring methods from Bayesian confirmation theory, I proceed in three steps: (1) I develop a framework for defining and comparing measures of causal strength; (2) I argue that no single measure can satisfy all natural constraints; (3) I prove two representation theorems for popular measures of causal strength: Pearl's causal effect measure and Eells' difference measure. In other words, I demonstrate these two measures can be derived from a set of plausible adequacy conditions. The paper concludes by sketching future research avenues.

1 Introduction

From Aristotle to the 21st century, causation is usually treated as a qualitative, all-or-nothing concept. Either C is a cause of E or it isn't. However, sometimes we have to make more nuanced causal judgments that involve a quantitative dimension: C is a more effective cause of E than C', the causal effect of C on E is twice as high as the effect of C', etc. This is especially important for purposes of prediction and evaluating experimental findings, e.g., quantifying effect sizes (e.g., Rubin, 1974; Rosenbaum and Rubin, 1983; Pearl, 2001). For instance, the extent to which a newly developed drug increases recovery rates affects the FDA's and EMA's decision to admit it to the market. The degree to which promotional activities cause high sales of a product affects the

*This is a half-baked draft which may nevertheless contain a couple of interesting results. Please do not cite without permission, and do not hesitate to contact me if you have comments/suggestions or spot mistakes.

†Contact information: Tilburg Center for Logic, Ethics and Philosophy of Science (TiLPS), Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. Email: j.sprenger@uvt.nl. Webpage: www.laeuferpaar.de

allocation of resources within a firm. The degree to which a car accident can be attributed to the behavior of an individual affects the amount of compensation that other parties receive. All these judgments tap onto the concept of *causal strength*, or equivalently, graded causation.

Whilst a huge amount of literature has been devoted to the qualitative question "When is C a cause of E?" (e.g., Hume, 1739; Suppes, 1970; Lewis, 1973; Mackie, 1974; Woodward, 2003), and the comparative question "Is C or C' a more effective cause of E?" starts to get explored as well (e.g., Chockler and Halpern, 2004; Halpern and Hitchcock, 2016), the quantitative question "What is the causal strength of C on E?" is relatively neglected, given the huge scope of actual and potential applications in science. There are proposals from different disciplines, such as psychology (Cheng, 1997), computer science (Pearl, 2000), statistics (Good, 1961a,b) and philosophy (Eells, 1991), but with the exception of Good's papers and the survey paper by Fitelson and Hitchcock (2011), no attempt is made at a unified theory of causal strength.

The purpose of this paper is to close this gap and to develop axiomatic foundations for probabilistic measures of causal strength: first, by developing a framework in which different measures can be defined and compared; second, by showing that not all natural constraints on causal strength can be jointly satisfied; and third, by proving two representation theorems for intuitive and much-discussed measures of causal strength. That is we show how the measures in question can be derived from a set of plausible adequacy conditions. This means that my paper is primarily descriptive, but to the extent that the adequacy conditions are plausible, it has normative implications for the choice of a measure of causal strength. It is also methodologically innovative in trying to develop a theory of causal strength that transfers formal methods which have been successful in the field of Bayesian confirmation theory. For a conference like FEW, bridging the causal strength and credence literature might be particularly exciting.

Laying foundations for a theory of causal strength also cuts across an important distinction in philosophical theories of causality: the distinction between type and token causation, or *actual* and *generic* causation. While reasoning about actual causation refers to concrete events and occurs in retrospective ("Was X or Y a cause of Z?", see e.g., Halpern and Pearl, 2005a,b), reasoning about generic causation refers to the causal relationship between two different events or event-types, regardless of their actual realization. One of the questions investigated by this paper is whether there should be different measures of causal strength, dependent on whether we ask a question about actual or about generic causation. (The answer is yes.)

In Section 2, I briefly motivate the choice of a probabilistic-interventionist framework for explicating causal strength. Then I provide a basic set of axioms from which a simple plurality result (we need more than one measure of causal strength) can be

derived. Section 3 introduces several adequacy criteria and shows representation theorems for measures of actual and generic causation, respectively. Section 4 discusses future research questions and concludes.

2 The Basic Axioms and the Plurality Result

This paper aims at an axiomatic theory of measures of causal strength for propositional variables, and a framework in which they can be compared. The theory is based on a probabilistic account of causal relevance (causes raise the probability of the effects), amended with the manipulability view of causation (Spirtes et al., 2000; Pearl, 2000; Woodward, 2003). On the interventionist account, C is a cause of E if and only if an *intervention* on C would cause a change of value in E, or change the probability that E takes a certain value (Woodward, 2012).[1]

An ideal intervention consists in forcing a variable C to take a certain value while breaking the influence that other variables may have on it. Pearl's notation for such an intervention is $do(C = x)$. Formally, this means "lifting C from the influence of the old functional mechanism and placing it under the influence of a new mechanism that sets the value $C = x$ while keeping all other mechanisms undisturbed" (Pearl, 2000, 70, notation changed). Imagine that we would like to study the effects of classroom light on whether students are awake or asleep. The intensity of classroom light depends on the settings of the audiovisual system. However, we may press the light switch manually, overruling the system settings, and then study the effects of our intervention on the students (e.g., they wake up from deep sleep). This way, we directly intervene on the light intensity and break the functional dependency on the audiovisual system settings.

A typical causal model in science contains a directed acyclical graph G that con-

[1]Outside formulae, notation in this paper distinguishes between propositional variables, which are printed in italics, and realizations of these variables, which are printed in regular roman script.

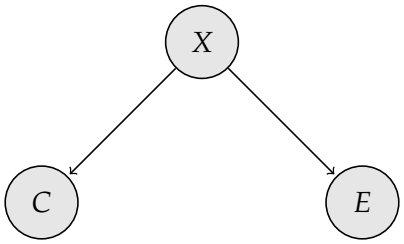

Figure 1: A typical common cause (conjunctive fork) structure. An intervention on C would disrupt the causal arrow leading into this variable from X and not have any effect on E.

sists of a set of vertices (=variables) and directed edges, together with a probability distribution over the variables in G.[2] In such a setting, the interventionist account naturally distinguishes genuinely causal relations between C and E from relations where both variables are correlated as a result of a common cause X. See Figure 1. When one intervenes on C, the causal arrow leading from X to C is broken and no effect on E occurs. On the probabilistic account, it is less straightforward to express this difference since common effects raise the probability of each other, too (e.g., Eells, 1991). While the probabilistic account describes causation in terms of statistical relevance, comparing $p(E|C)$ and $p(E|\neg C)$ in a set of relevant background contexts, the interventionist account focuses on probability of the effect conditional on an intervention on the cause, that is, $p(E|do(C))$. Both perspectives are combined in this paper.[3]

Now let L be a propositional language with closed sentences $\mathcal{L}$, and let $\mathcal{M}$ be the set of causal models whose variables are elements of $\mathcal{L}$. Each such model includes a joint probability distribution over its variables. This allows us to formulate our first axiom:

Axiom I (Domain): For any two propositions C, E $\in \mathcal{L}$ and a causal model $M \in \mathcal{M}$, the causal strength of C on E, $\eta(C, E)$, is a continuous real-valued function operating on a subset of $\mathcal{L}^2 \times \mathcal{M}$, namely the set

$$\mathcal{S} := \{\langle C, E, M\rangle \in \mathcal{L}^2 \times \mathcal{M} | \text{M contains C and E as variables}\} \tag{1}$$

Reference to a causal model in $\eta(C, E)$ is omitted for the sake of brevity. Moreover, define $\mathcal{C}$ as the set of measurable functions $\mathcal{S} \to \mathbb{R}$.

Implicitly, Axiom I asserts a strong claim: causal strength is a function of the joint probability distribution of the causal network together with information about the causal structure of M. This leaves out external factors such as typicality, normative expectations and defaults, which are of theoretical significance and have been shown to affect causal judgments in experimental settings (Knobe and Fraser, 2008; Hitchcock and Knobe, 2009; Halpern and Hitchcock, 2016). While this implies that my model does not capture all aspects of judgments of causal strength, there are many applications (e.g., quantifying effect size in science) where it is desirable to eliminate normative considerations, and to derive causal strength from observed relative frequencies. Moreover, my approach quantifies causal strength with respect to a single background context, sidestepping a substantial discussion in the field of probabilistic causation (e.g., Cartwright, 1979; Dupré, 1984; Eells, 1991).

[2] See Spirtes et al. (2000) for an introduction to causal reasoning with Bayesian nets.

[3] I am neutral on the interpretation of probability here: they can be subjective expectations (degrees of belief), actual or hypothetical frequencies, propensities, or Humean best-system-chances. Which interpretation is most appropriate will depend on the context of application, but our formal discussion is not touched by this.

The second axiom is purely technical and normalizes the range of measures of causal strength to the interval $[-1; 1]$.

Axiom II (Range): For all causal strength measures $\eta \in \mathcal{C}$, the range of η is the interval $[-1; 1]$.

The point of this axiom is that causal strength measures are easier to interpret and to compare if they all have the same scaling properties. The rich literature in Bayesian confirmation theory (e.g., Fitelson, 2001; Eells and Fitelson, 2002; Crupi, 2013) has shown that $[-1; 1]$ is an apt range. Since our project pursues a similar goal, we adopt the same technical constraints.

The next axiom has more philosophical substance. It connects the qualitative notion of being a cause to the causal strength measure. This move is borrowed from confirmation theory: There, E confirms H (in the qualitative sense) if and only if $c(E, H) > k$ for a suitable confirmation measure c and a critical value k. In that field, the bridge between the qualitative and quantitative concept has been very helpful at better understanding resolving longstanding problems of confirmation, such as the tacking paradoxes (e.g., Sprenger, 2016).

Axiom III (Qualitative-Quantitative Bridge Principle): For all measures of causal strength $\eta \in \mathcal{C}$,

- C is a (positive) cause of E if and only if $\eta(C, E) > 0$;
- C is a preventive cause of E if and only if $\eta(C, E) < 0$;
- C is causally irrelevant to E if and only if $\eta(C, E) = 0$.

In particular, this axiom clarifies that positive values of η denote a degree of causation, negative values denote degree of prevention, and zero is the neutral value. Axiom III could also be framed more generally with a variable k instead of the constant zero (see Crupi, 2013, for the case of confirmation), but like for Axiom II, I would like to keep things as simple and natural as possible. Note that Axiom III is specific to the case of propositional variables; extensions to other categorical and real-valued variables are not straightforward.

Now, I motivate another substantial constraint on η which also suits the interventionist approach well. It is motivated by a very old idea about causes, namely that causes make a *difference* to their effects. It has first been articulated by David Hume (1711–1776) in his famous description of two causally related objects: "if the first object had not been, the second never had existed" (Hume 1748/77).[4] This line of reasoning

[4]It should be kept in mind that Hume makes these remarks in the context of spelling out a regularity theory of causation, which is a bit at odds with the gist of the quote.

is later developed in the counterfactual account of David Lewis (1973, 1979), the probabilistic account of Patrick Suppes (1970) and Nancy Cartwright (1979), and exemplified in many cases of scientific inference, e.g., Randomized Controlled Trials (RCT). There, we would like to assess the efficacy of a drug and we divide the trial participants in two groups: one that receives the new drug, and one that receives the standard treatment or a placebo. The causal efficacy of the new drug to cure the disease is then a function of the divergence between the results in the treatment and the control group. At least *some* measure of causal strength should preserve this intuition. This motivates our next axiom:

Axiom IV (Probability Raising under Intervention) There is a measure of causal strength $\eta \in \mathcal{C}$ such that $\eta(C, E) > 0$ if and only if

$$p(E|do(C)) > p(E|do(\neg C))$$

Now consider a case of actual causation. Suzy throws stones at a bottle. What if she throws blindly? In comparison to a throw where she sees the bottle, throwing blindly lowers the probability that she hits the bottle. However, by accident her blind throw hits the wall behind the bottle and bounces back toward the bottle. The bottle shatters. Certainly, her blind throw was, in some sense, a cause of the shattering of the bottle. On the other hand, it lowered the probability of the effect. This is analogous to many cases in everyday life: making a technically sound decision at a card game such as poker or bridge will be right most of the time, but when the distribution of the cards is unlucky on a particular game, this decision may be the cause of a painful loss. Fortune has let you down. This motivates the following axiom:

Axiom V (Probability-Lowering Causes): There are causal models M with propositional variables C and E such that

$$p(E|do(C)) \leq p(E|do(\neg C))$$

and yet, C is a (positive) cause of E.

From the preceding axioms, it is easy to derive the **plurality of measures of causal strength** (all proofs in the appendix):

Theorem 1 *There is more than one measure of causal strength.*

This result is not particularly deep, and the proof is straightforward. It is reflected in the plurality of measures that Fitelson and Hitchcock (2011) list in their commendable survey article. The Suppes-Pearl measure, for example, violates Axiom IV and satisfies Axiom V, while the other measures satisfy Axiom IV and violate Axiom V. See

Suppes (1970); Pearl (2000)	$\eta(C,E) = p(E\|C)$
Eells (1991)	$\eta(C,E) = p(E\|C) - p(E\|\neg C)$
"Galton" (covariation)	$\eta(C,E) = 4p(C)\,p(\neg C)[p(E\|C) - p(E\|\neg C)]$
Lewis (1986)	$\eta(C,E) = \frac{p(E\|C) - p(E\|\neg C)}{p(E\|C) + p(E\|\neg C)}$
Cheng (1997)	$\eta(C,E) = \frac{p(E\|C) - p(E\|\neg C)}{1 - p(E\|C)}$
Good (1961a,b)	$\eta(C,E) = \frac{p(E\|C) - p(E\|\neg C)}{2 - p(E\|C) - p(E\|\neg C)}$

Table 1: Some prominent measures of causal strength. I follow the labels and rescalings by Fitelson and Hitchcock (2011).

Table 1. Yet, the plurality result deserves mention because it has not been noted explicitly in the literature on causal strength. For example, it has not been demonstrated that analyses of actual and generic causation may require different measures of causal strength.

The question is now which measures we should prefer, and this is a hard question whose answer will depend on the context of application. Nonetheless we will try to impose some constraints that are motivated by technical convenience or by plausible properties for causal relevance.

3 The Representation Theorems

The previous section has proposed that the causal strength of C for E depends on the causal model, that is, a directed acyclical graph in which C and E are included, with a probability distribution over the variables. We now make this intuition precise and demand that $\eta(C,E)$ be expressed as a function of the base rate of the cause and the probability of E under the relevant interventions on the cause.

Formality For two propositions $C, E \in \mathcal{L}$, the causal strength measure $\eta(C,E)$ is a real-valued function on $\mathcal{S} \subset (\mathcal{L}^2 \times \mathcal{M}) \to [-1;1]$, and there exists a measurable function $f : [0,1]^3 \to [-1;1]$ such that

$$\eta(C,E) = f(p(C), p(E|do(C)), p(E|do(\neg C)))$$

Formality thus takes up Axiom I and II from the previous section. The reader may ask why we are using the *do*-calculus in $p(E|do(C))$ and $p(E|do(\neg C))$ rather than the simple conditional probabilities $p(E|C)$ and $p(E|\neg C)$. The answer is that probability alone does not encapsulate information about causal directionality. Conditional probabilities alone do not specify whether E is a (positive) cause of C or the other way round, whether they have a common cause, etc. If $\eta(C, E)$ is to be more than a measure of statistical association or positive relevance between two variables, then these differences must be taken into account.

It is also notable that Formality is blind to mediator variables or multiple paths leading from C to E. Mediators will sometimes be latent variables and not be directly measurable. Therefore I keep the model simple and amalgamate the effects that C may have on E via different paths into one number (e.g., Dupré, 1984; Eells, 1991). Hoewever, this omission does not rule out a path-specific perspective. By appropriate conditionalization on other factors in the causal model, $\eta(C, E)$ can be used for calculating path-specific effects as well and comparing them to the net effect (cf., Pearl, 2001).

While Formality sketches the ground on which the different measures compete, the following adequacy conditions describe how they should rank different cause/effect pairs. Notably, not all of them will pull into the same direction. We start with the comparatively simple case of comparing two putative causes of an effect E. Suppose for example that we ask what is a stronger cause of headache (E): thinking hard about a difficult research problem (C_1) or going for a night of binge drinking (C_2)? In such cases, it is natural to answer that C_1 is more effective than C_2 if and only if C_1 makes E more expected than C_2:

Competing Causes I For propositions C_1, C_2, and $E \in \mathcal{L}$ and a causal model M with probability function $p(\cdot)$,

$$\eta(C_1, E) > \eta(C_2, E) \qquad \text{if and only if} \qquad p(E|do(C_1)) \; > \; p(E|do(C_2))$$

This requirement is analogous to Final Probability Incrementality in Bayesian confirmation theory (Crupi, 2013; Crupi et al., 2013). If it is found too restrictive, then the following weakening may be more palatable:

Competing Causes II

$$\eta(C_1, E) > \eta(C_2, E)$$

if and only if for some function $g : [0, 1]^2 \to \mathbb{R}$:

$$g(p(E|do(C_1)), p(E|do(\neg C_1))) > g(p(E|do(C_2)), p(E|do(\neg C_2)))$$

This condition is more liberal than Competing Causes I in demanding not more than that the base rates of C_1 and C_2 should not matter for ranking causal strength if the same effect is aimed at. Instead, we only look the degree to which intervening on C_1 and C_2 makes a difference for E. Note that Competing Causes II is silent on how the two conditional probabilities should be combined in determining causal strength.

The predictive function of causal strength suggests the following condition: C has positive (or at least not negative) causal strength for E if the intervention $do(\neg C)$ makes E impossible. In other words, if E cannot be explained but by C ($p(E|\neg C) = 0$), then $\eta(C, E) \geq 0$, and conversely for $\neg$C.

Inference to the Only Explanation If $p(E|do(C)) > 0$ and $p(E|do(\neg C)) = 0$ then $\eta(C, E) \geq 0$. Conversely, if $p(E|do(C)) = 0$ and $p(E|do(C)) > 0$, then $\eta(C, E) \leq 0$.

Rank	Team	Points	Team	Points
	after 36 out of 38 rounds		after 37 out of 38 rounds	
1	Roma	78	Inter	79
2	Inter	76	Roma	78
3	Juve	74	Juve	74

Table 2: A motivating example for Conditional Equivalence. Top of the Seria A after 36 and 37 out of 38 rounds, respectively.

Now we introduce a condition which is motivated by actual causation. Consider Table 2. Three teams in the Italian *Seria A*, AS Roma, FC Internazionale ("Inter"), and Juventus ("Juve") and, are competing for the *scudetto*, the national soccer championship. On the penultimate match day, Inter beats Juve in the *Derby d'Italia* while Roma loses to another team. Call this conjunction of events C. Let E_1 = Inter will win the championship and E_2 = Roma will be the runner-up. Given C, E_1 and E_2 are logically equivalent. (Juventus misses four and five points on both teams and cannot surpass them any more.) It is now very natural to claim that C has caused E_1 and E_2 to an equal degree. This intuition is stated in the following condition:

Conditional Equivalence If E_1 and E_2 are logically equivalent given C, then $\eta(C, E_1) = \eta(C, E_2)$.

Finally, a condition on how causal strength combines on a single path—see Figure 2. According to a plausible intuition, overall causal strength should be a function of individual causal strength. But which function $g : \mathbb{R}^2 \to \mathbb{R}$ should be chosen such that $\eta(C, E) = g(\eta(C, X), \eta(X, E))$? First of all, it appears natural that g is symmetric: the order of mediators in a chain does not matter. Second, it seems that the overall causal strength cannot be stronger than the weakest link in the chain: If C and X are

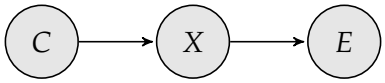

Figure 2: The Bayesian Network for causation along a single path.

almost independent, it doesn't matter how strongly X and E are correlated: the causal strength will be weak. Similarly, if both links are weak, the overall link will be even weaker. On the other hand, if the link is maximally strong (e.g., $\eta(C, X) = 1$), then the strength of the entire chain will just be the strength of the rest of the chain. See also Good (1961a, 311–312).

A very simple function that satisfies all these requirements is multiplication. Thus, we obtain as an adequacy criterion:

Multiplicativeness along Single Paths If the propositional variables C and E are connected via a single path with intermediate node X, then $\eta(C, E) = \eta(C, X) \cdot \eta(X, E)$.

As a corollary, we obtain that for a causal chain with multiple mediators, e.g., $C \to X_1 \to \ldots \to X_n \to E$,

$$\eta(C, E) = \eta(C, X_1) \cdot \eta(X_1, X_2) \cdot \ldots \cdot \eta(X_{n-1}, X_n) \cdot \eta(X_n, E)$$

This concludes the exposition of adequacy criteria. Note that the point of this section is not to find a measure of causal strength that satisfies all adequacy conditions. Actually, one can even show that no such measure exists. However, we can use subsets of the conditions for characterizing different measures up to ordinal equivalence. Two measures η and η' are called *ordinally equivalent* if and only if they agree in their *rankings* of causal strength, that is, $\eta(C, E) > \eta(C', E')$ if and only if $\eta'(C, E) > \eta'(C', E')$.

I now state two representation theorems pertaining to these adequacy conditions.

Theorem 2 *All measures of causal strength that satisfy Formality, Competing Causes I and Conditional Equivalence are ordinally equivalent to*

$$\eta_*(C, E) = p(E|do(C))$$

Pearl (2000, 70) calls $\eta_*(C, E) = p(E|do(C))$ the "causal effect" of C on E. This measure fits quite well with cases of actual causation where we are asked to rank causes of an event according to the degree that they produced E or were responsible for E. For instance, should a car accident (E) be attributed to driving a bit too fast (C_1) or to ignoring a red traffic light (C_2)? Although both causes describe the violation of a norm, one of them has a much higher tendency to cause an accident, and $p(E|do(C))$

seems to be a good guide for ranking the causes. Note, however, that η_* violates Axiom III and Axiom IV from the previous section: it is not suitable to distinguish between (positive) causal relevance, causal prevention, and causal irrelevance. It also violates Multiplicity along Single Paths and Inference to the Only Explanation.

We now prove a representation theorem for a measure that is based on statistical relevance, and that agrees with Axiom I-IV.

Theorem 3 *All measures of causal strength that satisfy Formality, Inference to the Only Explanation, Competing Causes II, and Multiplicativity along Single Paths are ordinally equivalent to*

$$\eta_d(C, E) = p(E|do(C)) - p(E|do(\neg C))$$

This is a simple and intuitive quantity that measures the causal strength of C for E by comparing the effect that different interventions on C have on E. It possesses the *sine qua non* property that two effects in a conjunctive fork (e.g., $E_1 \leftarrow C \rightarrow E_2$) do not cause each other. It is also straightforwardly applicable in statistical inference where it is used to quantify effect size for categorical variables under an intervention on C. For example, in clinical trials, $\eta_d(C, E)$ is called a measure of Absolute Risk Reduction, or ARR.

We conclude this section by stating some notable properties of η_d. First, it can be rewritten as

$$\begin{aligned} \eta_d(C, E) &= p(E|do(C)) - p(E|do(\neg C)) \\ &= p(\neg E|do(\neg C)) + p(E|do(C)) - 1 \end{aligned}$$

Modulo subtraction of a constant, $\eta_d(C, E)$ is a sum of two quantities that have been called causal/explanatory necessity and causal/explanatory sufficiency by Hempel (1965) and Pearl (2000). The names are natural: $p(\neg E|do(\neg C))$ indicates to what extent C was *necessary* for producing E (what would have happened if E had not occurred?), and $p(E|do(C))$ indicates to what extent the presence of C is *sufficient* for producing E. $\eta_d(C, E)$ combines these two plausible ways of thinking about causal strength in an intuitive manner.

While this property may be regarded as superficial, the following one is more profound. Consider the proposition E_1 that a certain real-valued quantity E falls into the interval $[e_1^-, e_1^+]$ and the proposition E_2 that E has values in $[e_2^-, e_2^+]$. Obviously, these two variables are mutually exclusive. But what is the degree to which C causes E_1 *or* E_2 (that is, $E \in [e_1^-, e_1^+] \cup [e_2^-, e_2^+]$)? This question can be answered in general:

Corollary 1 For C, E_1 and $E_2 \in \mathcal{L}$,

$$\eta_d(C, E_1 \vee E_2) = \eta_d(C, E_1) + \eta_d(C, E_2) - \eta_d(C, E_1 \wedge E_2) \tag{2}$$

and in particular, if E_1 and E_2 are mutually exclusive (that is, if $\neg(E_1 \wedge E_2)$ is a theorem), then the above equation reduces to

$$\eta_d(C, E_1 \vee E_2) = \eta_d(C, E_1) + \eta_d(C, E_2)$$

and we can also formulate the following necessary and sufficient condition on rankings of causal strength:

$$\eta_d(C, E_1 \vee E_2) > \eta_d(C, E_1) \qquad \text{iff} \qquad \eta_d(C, E_2) > 0,$$

and vice versa with E_1 and E_2 reversed.

The proof is straightforward and left as an exercise. This means that the degree to which a (mutually exclusive) disjunction of effects is caused is the sum of the individual degrees of causation. In particular, causal strength is enlarged by disjunctively tacking further effects if and only if each of these effects is itself caused to a positive degree. Notably, it would be possible to characterize $\eta_d(C, E)$ up to ordinal equivalence by conjoining (2) with Competing Causes I (cf. Crupi, 2013; Crupi et al., 2013).

Third and last, $\eta_d(C, E)$ satisfies a natural symmetry constraint proposed by Fitelson and Hitchcock (2011), namely that the degree to which C prevents E (=the degree to which C causes ¬E) is the negative of the degree to which C causes E:

$$-\eta(C, E) = \eta(C, \neg E) \qquad \text{(Causation-Prevention Continuity or CPC)}$$

This concludes our discussion of the adequacy conditions and the measures which can be used to derive them.

4 Discussion

This paper has provided axiomatic foundations for a probabilistic theory of causal strength, proceeding toward a more systematic investigation of that topic. It synthesizes ideas from the manipulability/interventionist view of causation and the probabilistic relevance view of causation in developing a measure of causal strength. The methods for characterizing the various measures are partly transferred from Bayesian confirmation theory.

After an introduction of the conceptual and mathematical framework, the paper has essentially shown two results. First, intuitions about measures of causal strength pull into different directions, making it difficult to come up with one true measure of causal strength. Second, by means of a list of plausible adequacy criteria, one can characterize two intuitive measures of causal strength: $\eta_*(C, E) = p(E|do(C))$ for matters of actual causation and causal attribution, $\eta_d(C, E) = p(E|do(C))$ for matters of

Measure	Property						
	FORM	CC1	CC2	CE	IOE	MUL	CPC
Suppes/Pearl	yes	yes	yes	yes	no	no	no
Eells	yes	no	yes	no	yes	yes	yes
Cheng	yes	no	yes	no	yes	no	no
Galton	yes	no	no	no	yes	no	yes
Lewis	yes	no	yes	no	yes	no	no
Good	yes	yes	no	no	yes	no	no

Table 3: A classification of different measures of causal strength according to the adequacy conditions that they satisfy. FORM = Formality, CC1+2 = Competing Causes I+II, CE = Conditional Equivalence, IOE = Inference to the Only Explanation, MUL = Multiplicativity along Single Paths, CPC = Causation-Prevention Continuity.

generic causation and predicting the effect of an intervention. As stated in the previous section, these measures are not only conceptually appealing, but they also have plenty of applications in scientific and everyday reasoning. For instance, Stegenga (2015) has argued that η_d is superior to competing causal effect measures in medical decision-making, e.g., the Lewis measure, both from epistemological and pragmatic points of view. All these arguments make a good case for treating η_* and η_d as default measures for causal strength, while leaving open that other measures may be more appropriate for specific applications.

What remains to do? First of all, we may aim at generalizing the framework from propositional variables to categorical and real-valued variables. Indeed, many measures of effect size for real-valued variables, such as Cohen's d or Glass's Δ, are based on the difference of group means, and η_d might be extended naturally into this direction.

Second, it would be desirable to axiomatize other measures of causal strength along similar lines. That would help us to nail down the differences between them in a formally precise way, and help to argue for or against specific measures. Table 3 already displays which measure in Table 1 satisfies which adequacy conditions. It is notable that few measures satisfy Multiplicativity along Single Paths and the Causation-Prevention Continuity, although these are eminently sensible properties.

Third, the properties of η_d in complicated networks (e.g., more than one path linking C and E) have not been investigated. Is it possible to show, for example, how degrees of causation along different paths can be combined in an overall assessment of causal strength, e.g., similar to Theorem 3 in Pearl (2001)?

Fourth, this work can be connected to information-theoretic approaches to *causal specificity* (Weber, 2006; Waters, 2007; Korb et al., 2011; Griffiths et al., 2015). The more narrow the range of effects that an intervention can produce, the more specific the

cause is to the effect. How does this concept relate to causal strength and causal effect and to what extent can both research programs learn from each other?

Fifth, I would like to apply this theory to canonical examples in the causation literature and to explore whether this understanding of causal strength squares well with the significance of normality and norms in causal reasoning (Knobe and Fraser, 2008; Hitchcock and Knobe, 2009).

These are all open and exciting questions, and I guess it is not difficult to come up with others. The present paper is just a beginning. I hope, however, that the results presented herein are promising enough to motivate a further pursuit of an axiomatic theory of causal strength.

Proofs of the Theorems

Proof of Theorem 1 (Plurality of Measures of Graded Causation)

We have to show that Axiom III, IV and V are not compatible. From Axiom V, we know that there is a causal model with a probability-lowering cause $p(E|do(C)) < p(E|do(\neg C))$. By Axiom III, $\eta(C,E) > 0$ for all causal strength measures $\eta \in \mathcal{C}$. This conflicts straightforwardly with Axiom IV, which asserts that there is at least one measure of causal strength that tracks probability-raising: $\eta(C,E) > 0$ if and only if $p(E|do(C)) > p(E|do(\neg C))$. Contradiction. □

Proof of Theorem 2 (Representation Theorem for Graded Actual Causation)

The proof relies on a recent result by Michael Schippers (2016) in the field of confirmation theory. Schippers demonstrates that the following three conditions are necessary and sufficient to characterize the posterior probability $c^*(E,H) := p(H|E)$ as a measure of degree of confirmation, up to ordinal equivalence.

Formality (Confirmation) There is a measurable function $f' : [0,1]^3 \to \mathbb{R}$ such that for any $h, e \in \mathfrak{L}$ with probability distribution $p(\cdot)$, $c(E,H) = f'(p(E), p(H), p(H \wedge E))$.

Final Probability Incrementality For any sentences H, E_1, and $E_2 \in \mathfrak{L}$ with probability measure $p(\cdot)$,

$$c(E_1,H) > c(E_2,H) \qquad \text{if and only if} \qquad p(H|E_1) > p(H|E_2).$$

Local Equivalence If H_1 and H_2 are logically equivalent given E, then $c(E,H_1) = c(E,H_2)$.

It is easy to see that Final Probability Incrementality translates into Competing Causes I when the pair $(H, E_{1,2})$ is mapped to $(E, C_{1,2})$:

$$\eta(C_1, E) > \eta(C_2, E) \qquad \text{if and only if} \qquad p(E|C_1) \; > \; p(E|C_2)$$

The same is true for Local Equivalence: with $(H_{1,2}, E)$ mapped to $(E_{1,2}, C)$, it postulates that if E_1 and E_2 are logically equivalent given C, then $\eta(C, E_1) = \eta(C, E_2)$. This is just the same as Conditional Equivalence.

Thus it remains to show that Formality (Graded Causation) can be transformed into Formality (Confirmation) by a suitable change of variables. We already know that there exists a $f : [0,1]^3 \to \mathbb{R}$ such that $\eta(C, E) = f(p(C), p(E|do(C)), p(E|do(\neg C))$. Since we only want to characterize f mathematically, we restrict ourselves to the case where E is among the descendants of C and they share no common causes. We also assume that $p(C) \in (0, 1)$. This allows us to write

$$p(E \wedge C) = p(C)p(E|do(C)) \quad p(E) = p(C)p(E|do(C)) + (1 - p(C))p(E|do(\neg C))$$

which we can transform into the equations

$$p(E|do(C)) = \frac{p(E \wedge C)}{p(C)} \qquad p(E|do(\neg C)) = \frac{p(E) - p(C)p(E|do(C))}{1 - p(C)} \tag{3}$$

Hence, we can write $p(E|do(C))$ and $p(E|do(\neg C))$ as functions of $p(C)$, $p(E)$ and $p(E \wedge C)$. In other words, there is a function $f'(p(C), p(E), p(C \wedge E))$ that characterizes $\eta(C, E)$, namely

$$\begin{aligned} f'(p(C), p(E), p(C \wedge E)) &:= f\left(p(C), \frac{p(E \wedge C)}{p(C)}, \frac{p(E) - p(C)p(E|do(C))}{1 - p(C)}\right) \\ &= f(p(C), p(E|do(C)), p(E|do(\neg C)) \\ &= \eta(C, E) \end{aligned}$$

f' is continuous because f and the functions in Equation (3) are. Thus we can extend f' canonically to the set $\{p(C) \in \{0, 1\}\}$. Hence we can invoke Schippers' theorem which shows that $\eta(C, E) = p(E|C)$ up to ordinal equivalence. □

Proof of Theorem 3 (Representation Theorem for Graded Generic Causation)

The proof of this representation theorem proceeds in several steps. First, we will show that $\eta(C, E) = f(p(C), p(E|do(C)), p(E|do(\neg C))$ does not depend on $p(C)$.

The proof of this first claim proceeds by contradiction. Consider that there are real numbers $x_1, x_2, y, z \in [0, 1]$ such that $f(x_1, y, z) \neq f(x_2, y, z)$. Then choose E,

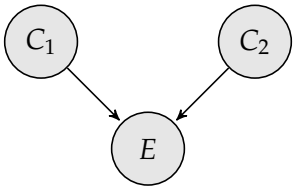

Figure 3: A classical collider/joint effect structure in a causal net.

C_1 and C_2 such that E is a joint effect of C_1 and C_2 with $x_1 = p(C_1)$, $x_2 = p(C_2)$, $y = p(E|do(C_1)) = p(E|do(C_2))$, $z = p(E|do(\neg C_1)) = p(E|do(\neg C_2))$. In this case, Competing Causes II tells us that $\eta(C_1, E) = \eta(C_2, E)$. However, on the other hand, we also know

$$\eta(C_1, E) = f(x_1, y, z) \neq f(x_2, y, z) = \eta(C_2, E)$$

This leads to a straightforward contradiction. Hence, from now on we focus on the function $g : [0, 1]^2 \to \mathbb{R}$ such that $\eta(C, E) = g(p(E|do(C)), p(E|do(\neg C))$.

The second step of the proof consists in deriving the equality

$$g(\alpha, \bar{\alpha}) \cdot g(\beta, \bar{\beta}) = g(\alpha\beta + (1 - \alpha)\bar{\beta}, \bar{\alpha}\beta + (1 - \bar{\alpha})\bar{\beta}) \tag{4}$$

To this end, recall the Bayesian network from the main paper. It is reproduced in Figure 4. Again, for the purpose of investigating the formal properties of g, we can focus on those cases where $p(E|\pm C)$ and $p(E|\pm C)$ agree.

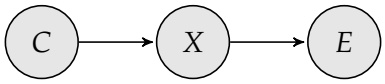

Figure 4: The Bayesian Network for causation along a single path.

We know by Multiciplity along Single Paths that

$$\begin{aligned} \eta(C, E) &= \eta(C, X) \cdot \eta(X, E) \\ &= g(p(X|do(C)), p(X|do(\neg C))) \cdot g(p(E|do(X)), p(E|do(\neg X))) \\ &= g(p(X|C), p(X|\neg C)) \cdot g(p(E|X), p(E|\neg X)) \end{aligned}$$

and at the same time,

$$\begin{aligned} \eta(C, E) &= g(p(E|do(C)), p(E|do(\neg C))) \\ &= g\left(\sum_{\pm X} p(X|C)p(E|C, X), \sum_{\pm X} p(X|\neg C)p(E|\neg C, X)\right) \end{aligned}$$

Combining both equations yields

$$g(p(X|C), p(X|\neg C)) \cdot g(p(E|X), p(E|\neg X)) \quad = \quad g\left(\sum_{\pm X} p(X|C)p(E|C,X), \sum_{\pm X} p(X|\neg C)p(E|\neg C,X)\right)$$

With the variable settings

$$\begin{aligned} \alpha &= p(X|C) & \beta &= p(E|X) \\ \bar{\alpha} &= p(X|\neg C) & \bar{\beta} &= p(E|\neg X) \end{aligned}$$

equation (4) follows immediately.

Third, we are going to show that

$$g(x, y) = g(x - y, 0) \tag{5}$$

To this end, we first note a couple of facts about g:[5]

Fact 1 $g(\alpha, 0)g(\beta, 0) = g(\alpha\beta, 0)$. This follows immediately from equation (4) with $\bar{\alpha} = \bar{\beta} = 0$.

Fact 2 $g(1, 0) = 1$. With $\beta = 1$, the previous fact entails that $g(\alpha, 0)g(1, 0) = g(\alpha, 0)$. Hence, either $g(\alpha, 0) \equiv 0$ for all values of α (which would trivialize g) or $g(1, 0) = 1$.

Fact 3 $g(0, 1) = -1$. Fact 1 entails (with $\alpha = \beta = 0$, $\bar{\alpha} = \bar{\beta} = 1$) that $g(0, 1) \cdot g(0, 1) = g(1, 0) = 1$. Hence, either $g(0, 1) = -1$ or $g(0, 1) = 1$. If the latter were the case, then g would take positive values although $p(E|do(C)) = 0$ and $p(E|do(\neg C)) > 0$, in violation of Inference to the Only Explanation. Thus, $g(0, 1) = -1$.

Fact 4 $g(-1, 0) = -1$. By Fact 1, $g(-1, 0) \cdot g(-1, 0) = g(1, 0) = 1$. Then we apply the same reasoning as in the proof of Fact 3.

Fact 5 $g(0, 1) \cdot g(\beta, \bar{\beta}) = g(\bar{\beta}, \beta)$. Follows immediately from equation (4) with $\alpha = 0$, $\bar{\alpha} = 1$.

These facts will allow us to derive Equation (5). Note that (5) is trivial if $y = 0$. So we can restrict ourselves to the case that $y > 0$. We choose the variable settings

$$\begin{aligned} \alpha &= \frac{y - x}{y} & \beta &= 0 \\ \bar{\alpha} &= 0 & \bar{\beta} &= y \end{aligned}$$

[5] In the proof, negative arguments of g figure. This may look problematic, but it isn't. We just show that any $g(\cdot, \cdot)$ that satisfies Equation (4) on $[0, 1]^2$ has an extension to a function on $\mathbb{R}^2$ that satisfies certain properties, which can in turn be used for saying something about the behavior of g on $[0, 1]^2$.

Then we obtain by means of Equation (4) and the previously proven facts

$$\begin{aligned} g(x,y) &= g((y-x)/y,0) \cdot g(0,y) \\ &= g(y-x,0) \cdot g(1/y,0) \cdot g(0,y) \qquad \text{(Fact 1)} \\ &= g(y-x,0) \cdot g(1/y,0) \cdot g(y,0) \cdot g(0,1) \qquad \text{(Fact 5)} \\ &= g(y-x,0) \cdot g(1,0) \cdot g(-1,0) \qquad \text{(Fact 1+3+4)} \\ &= g(x-y,0) \qquad \text{(Fact 1+2)} \end{aligned}$$

This implies

$$\eta(C,E) = g(p(E|do(C)), p(E|do(\neg C))) = g(p(E|do(C)) - p(E|do(\neg C)), 0)$$

Hence, $\eta(C,E)$ is a function of $p(E|do(C)) - p(E|do(\neg C))$ only. It is easy to see that this function must be monotonic, that is, g is monotonically increasing in its first argument. Otherwise there would be $x, y \in [0,1]$ with $x > y$ and $g(x,0) < g(y,0)$. In that case, application of Equation (5) and Inference to the Only Explanation yields

$$0 > g(x,0) - g(y,0) = g(x-y,0) \geq 0$$

and a contradiction results. This concludes the proof of the Theorem. □

References

Cartwright, N. (1979). Causal Laws and Effective Strategies. *Noûs*, 13(4):419–437.

Cheng, P. W. (1997). From Covariation to Causation: A Causal Power Theory. *Psychological Review*, 104(2):367–405.

Chockler, H. and Halpern, J. Y. (2004). Responsibility and blame: A structural-model approach. *Journal of Artificial Intelligence Research*, 22:93–115.

Crupi, V. (2013). Confirmation. *The Stanford Encyclopedia of Philosophy*.

Crupi, V., Chater, N., and Tentori, K. (2013). New Axioms for Probability and Likelihood Ratio Measures. *British Journal for the Philosophy of Science*, 64(1):189–204.

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

Eells, E. (1991). *Probabilistic causality*. Cambridge University Press, Cambridge.

Eells, E. and Fitelson, B. (2002). Symmetries and Asymmetries in Evidential Support. *Philosophical Studies*, 107(2):129–142.

Fitelson, B. (2001). *Studies in Bayesian Confirmation Theory*. PhD thesis, University of Wisconsin - Madison.

Fitelson, B. and Hitchcock, C. (2011). Probabilistic Measures of Causal Strength. In Illari, P. M., Russo, F., and Williamson, J., editors, *Causality in the Sciences*, pages 600–627. Oxford University Press, Oxford.

Good, I. J. (1961a). A Causal Calculus (I). *British Journal for the Philosophy of Science*, 11(44):305–318.

Good, I. J. (1961b). A Causal Calculus (II). *British Journal for the Philosophy of Science*, 12(45):43–51.

Griffiths, P. E., Pocheville, A., Calcott, B., Stotz, K., Kim, H., and Knight, R. (2015). Measuring Causal Specificity. *Philosophy of Science*, 82(4):529–555.

Halpern, J. Y. and Hitchcock, C. (2016). Graded causation and defaults. *The British Journal for the Philosophy of Science*.

Halpern, J. Y. and Pearl, J. (2005a). Causes and Explanations: A Structural-Model Approach. Part I: Causes. *British Journal for the Philosophy of Science*, 56(4):843–887.

Halpern, J. Y. and Pearl, J. (2005b). Causes and Explanations: A Structural-Model Approach. Part II: Explanations. *British Journal for the Philosophy of Science*, 56(4):889–911.

Hempel, C. G. (1965). Aspects of Scientific Explanation. In *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*, pages 331–496. Free Press, New York.

Hitchcock, C. and Knobe, J. (2009). Cause and norm. *The Journal of Philosophy*, 106(11):587–612.

Hume, D. (1739). *A Treatise of Human Nature*. Clarendon Press, Oxford.

Knobe, J. and Fraser, B. (2008). Causal judgment and moral judgment: Two experiments. *Moral psychology*, 2:441–448.

Korb, K. B., Nyberg, E. P., and Hope, L. (2011). A new causal power theory. In Illari, P., Russo, F., and Williamson, J., editors, *Causality in the Sciences*, pages 628–652. Oxford University Press, Oxford.

Lewis, D. (1973). Causation. *Journal of Philosophy*, 70:556–567.

Lewis, D. (1979). Counterfactual Dependence and Time's Arrow. *Noûs*, 13:455–476.

Lewis, D. (1986). *Philosophical Papers, Volume 2*. Oxford University Press, Oxford.

Mackie, J. L. (1974). *The Cement of the Universe: a study in Causation*. Clarendon Press, Oxford.

Pearl, J. (2000). *Causality*. Cambridge University Press, Cambridge.

Pearl, J. (2001). Direct and Indirect Effects. In *Proceedings of the Seventeenth Conference on Uncertainy in Arifical Intelligence*, pages 411–420.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology*, 66(5):688–701.

Schippers, M. (2016). A representation theorem for absolute confirmation.

Spirtes, P., Glymour, C. N., and Scheines, R. (2000). *Causation, prediction, and search*. Springer, New York.

Sprenger, J. (2016). Confirmation and Induction. In Humphreys, P. W., editor, *Handbook of Philosophy of Probability*. Oxford University Press, Oxford.

Stegenga, J. (2015). Measuring effectiveness. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 54:62–71.

Suppes, P. (1970). *A Probabilistic Theory of Causality*. North-Holland, Amsterdam.

Waters, C. K. (2007). Causes That Make a Difference. *Journal of Philosophy*, 104(11):551–579.

Weber, M. (2006). The Central Dogma as a Thesis of Causal Specificity. *History and Philosophy of the Life Sciences*, 28:595–609.

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

Woodward, J. F. (2012). Causation and Manipulability. *Stanford Encyclopedia of Philosophy*.