Foundations of a Probabilistic Theory of Causal Strength

Jan Sprenger*

November 18, 2016

Abstract

This paper develops axiomatic foundations for a probabilistic theory of causal strength. I proceed in three steps: First, I motivate the choice of causal Bayes nets as a framework for defining and comparing measures of causal strength. Second, I prove several representation theorems for probabilistic measures of causal strength—that is, I demonstrate how these measures can be derived from a set of plausible adequacy conditions. Third, I compare these measures on the basis of their characteristic properties, including an application to quantifying causal effect in medicine. Finally, I use the above results to argue for a specific measure of causal strength and I outline future research avenues.

*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

Contents

1 Introduction **3**

2 The Framework: Causal Bayes Nets **7**

2.1 Probabilistic and Interventionist Accounts of Causation 7

2.2 Explicating the Determinants of Causal Strength 10

3 The Representation Theorems **14**

3.1 General Adequacy Conditions 14

3.2 Causal Production and the Pearl Measure 17

3.3 Separability of Effects and the Probability-Raise Measure 19

3.4 Multiplicativity on Single Paths and the Difference Measure 21

3.5 No Dilution for Irrelevant Effects and Probability Ratio Measures 23

3.6 Conjunctive Closure and the Logarithmic Ratio Measure 26

4 Application: Quantifying Causal Effect in Medicine **28**

5 Discussion **30**

A Proofs of the Theorems **35**

1 Introduction

Causation is a central concept in human cognition. Knowledge of causal relationships enables us to make predictions, to explain phenomena, and to understand complex systems. Decisions are taken according to the effects which they are supposed to bring about. Actions are evaluated according to whether or not they caused an event.

Since the days of Aristotle, causation has been treated primarily as a qualitative, all-or-nothing concept. 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). 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, 2015). By contrast, the quantitative question "What is the strength of the causal relationship between C and E?" is relatively neglected. This is surprising given the huge scope of applications in science, such as prediction, control and manipulation, where causal judgments regularly 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', and so on (e.g., Rubin, 1974; Rosenbaum and Rubin, 1983; Pearl, 2001). For instance, regression coefficients are routinely interpreted as expressing the causal strength of the predictor variable for the dependent variable. Regulatory medical bodies only admit those drugs that have a substantial causal effect on recovery. Cognitive psychologists use the concept of causal power to analyze human reasoning (Cheng, 1997; Waldmann and Hagmayer, 2001; Sloman and Lagnado, 2015). In tort law, the causal contributions of two actions to a damage (e.g., a car accident) have to be weighed in order to determine individual liability (Rizzo and Arnold, 1980; Kaiserman, 2016b). All these judgments tap onto the concept of **causal strength**, and the size of a causal effect.

Proposals for explicating this concept are rare and spread over different disciplines, each with their own motivation and intended context of application. This includes cognitive psychology (Cheng, 1997), computer science and machine learning (Pearl, 2000;

Korb et al., 2011), statistics (Good, 1961a,b; Holland, 1986; Cohen, 1988), epidemiology and clinical medicine (Davies et al., 1998), philosophy of science (Suppes, 1970; Eells, 1991), political philosophy and social choice theory (Braham and van Hees, 2009), and legal theory (Rizzo and Arnold, 1980; Hart and Honoré, 1985). Although most of these disciplines use a common formalism—probability theory—to express causal strength, few axiomatic, principle-based explications are offered and normative comparisons of different measures are rare (though see the survey paper by Fitelson and Hitchcock, 2011). That is the gap that this paper tries to fill.

Group/Outcome	Effect	No Effect	Total Number
Treatment 1	A=20	B=100	A+B=120
Control 1	C=10	D=110	C+D=120
Treatment 2	A=45	B=75	A+B=120
Control 2	C=30	D=90	C+D=120

Table 1: The result of two clinical trials where the efficacy of a new migraine treatment is compared to a control group. How should the causal strength of the treatment, as opposed to the control, be quantified?

Measures of causal strength can differ substantially. The following example illustrates the differences and their practical implications. Consider a Randomized Controlled Trial (RCT) where we administer a new migraine drug to a group of patients and compare this treatment group to a control group which receives the standard drug. Causal effect is measured on a binary scale: was the pain relieved or not? Suppose the results are described by Table 1. In the epidemiological literature, several measures have been proposed to measure the strength of such an effect (Davies et al., 1998; Deeks, 1998; Sistrom and Garvan, 2004; King et al., 2012):

Relative Risk (RR) The ratio of the relative frequencies of an effect in both groups.

$$\mathrm{RR} = \frac{A/(A+B)}{C/(C+D)}$$

Absolute Risk Reduction (ARR) The difference between the relative frequencies of

an effect in both groups.

$$\text{ARR} = \frac{A}{A+B} - \frac{C}{C+D}$$

Relative Risk Reduction (RRR) The difference between the relative frequencies of an effect in both groups, relative to the frequency in the control group.

$$\text{RRR} = \frac{A/(A+B) - C/(C+D)}{C/(C+D)}$$

Table 1 presents the outcomes of the trial (Treatment 1, Control 1). The relative risk is RR $= 2$, meaning that the treatment halves the frequency of pain in the affected population. The result looks similar for relative risk reduction RRR $= 1$, which amounts to a 100% increase of recovery in the treatment group. However, the absolute risk reduction ARR $= 0.083$ tells a less enthusiastic story: only for 8,3% of the trial population, the new treatment makes a difference. Since such measurements of causal strength directly affect clinical decision-making, the question which measure should be preferred is vital and urgent (e.g., Stegenga, 2015; Sprenger and Stegenga, 2016).

To press this point even more, suppose that a second migraine drug is tested in a different population. The design is again a Randomized Controlled Trial and the data are given in the rows (Treatment 2, Control 2) in Table 1. We observe RR $= 1.5$, RRR $= 0.5$ and ARR $= 0.15$. That is, while an evaluation in terms of RR and RRR declares Drug 1 as the stronger cause of pain relief, an evaluation in terms of ARR would lead to the conclusion that the Drug 2 is the much stronger cause of pain relief. This poses an immediate practical dilemma for the doctor who is advising a patient suffering from migraine. Which drug is more effective? Which one should she prescribe? Similar examples occur in other disciplines whenever the causal strength of two different interventions is compared.

The challenge for a philosophical theory of causal strength is to characterize the various measures and to weigh the reasons for preferring one of them over another.

In this paper, I develop axiomatic foundations for measures of causal strength. First, I motivate the choice of causal Bayes nets as a framework where different measures can be embedded and compared. The relata of the causal strength relation are conceptualized as instantiations of binary variables (e.g., propositions). Second, I derive representation theorems for various measures of causal strength, that is, theorems that characterize a measure of causal strength in terms of a set of adequacy conditions. Third, I compare and discuss these measures with a view towards applications: Under which conditions are they invariant? What are beneficial and what are problematic properties? To the extent that the proposed adequacy conditions are found compelling, the technical results have normative implications for the choice of a measure of causal strength. Indeed, I will make a case for a particular measure as opposed to working with a plurality of causal strength measures. This distinguishes my results from investigations in formal epistemology where pluralism is the default position regarding measures of confirmation and explanatory power (e.g., Schupbach and Sprenger, 2011; Crupi and Tentori, 2012, 2013; Crupi et al., 2013).

The remainder of this paper is structured as follows: Section 2 motivates the choice of causal Bayes nets as a framework for explicating causal strength and specifies the formal structure of a causal strength measure. Section 3 contains the core of the paper: the representation theorems for the various causal strength measures in terms of a set of axioms or adequacy conditions. The methods of that section are partially taken from confirmation theory and related areas of formal epistemology—an interesting case of cross-fertilization in philosophy of science. Section 4 returns to the above medical science example and applies the formal results of Section 3, while Section 5 discusses future research questions and concludes. The appendix contains the proofs of the theorems.

2 The Framework: Causal Bayes Nets

2.1 Probabilistic and Interventionist Accounts of Causation

Causes matter for their effects. This thought has already 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). This line of reasoning has been developed later in counterfactual accounts of causation (Lewis, 1973, 1979).

However, causes do not always necessitate their effects. We classify smoking as a cause of lung cancer although not every regular smoker will eventually suffer from lung cancer. The same is true in other fields of science, e.g., when we conduct psychological experiments or make economic policy decisions: interventions increase the frequency of a particular response, but they do not guarantee it. Therefore, causal relevance is often explicated as probability-raising, especially when reasoning at the population level.

On the **probabilistic account of causation**, C is a cause of E if and only if C raises the probability that E occurs (e.g. Reichenbach, 1956; Suppes, 1970; Cartwright, 1979; Eells, 1991). This account captures the intuition that many causes make a difference to their effects without necessitating them. Moreover, this account squares well with many cases of scientific inference, like RCTs, where we compare the effect of two different interventions. In such cases (e.g., the data from Table 1), it seems that the stronger the divergence between the two groups, the greater the causal strength. Indeed, probabilistic measures of causal strength, with their natural link to observed relative frequencies, nicely square with an account of causation where causes raise the occurrence rate of an effect.

However, a purely probabilistic account struggles with various aspects of causal reasoning. Pearl (2000, 2011) objects that causal claims go beyond the pure associational level that is encoded in probability distributions; they express how the world would change in response to interventions. Hence,

> "[e]very claim invoking causal concepts must rely on some premises that invoke such concepts; it cannot be inferred from, or even defined in terms of statistical associations alone." (Pearl, 2011, 700)

The classical example is a case where cause and effect are correlated, but independent conditional on a common cause. In the Netherlands, a high crime rate (E) has recently been found to be correlated with a high percentage of migrants living in the affected neighborhoods (C). In particular, $p(\mathrm{E}|\mathrm{C}) > p(\mathrm{E}|\neg\mathrm{C})$. This suggests that a high rate of migrants in a neighborhood leads to more crimes being committed. However, the correlation could be explained by a common cause: the low socio-economic status of these neighborhoods (Jensma, 2014). Call this variable X. Conditional on the various levels of average income in a neighborhood, there was no correlation between the crime rate and the number of migrants: $p(\mathrm{E}|\mathrm{C}, X = x) \approx p(\mathrm{E}|\neg\mathrm{C}, X = x)$. Figure 1 gives a graphical representation. The naïve probabilistic account gets this wrong and judges the number of migrants to be a cause of the high crime rate.

To solve this problem, contextual unanimity has been demanded: the putative cause has to raise the probability of the effect in every background context (e.g., Cartwright, 1979). However, such a condition is very strict: causal relations vanish as soon as there is a single background context where the probability is lowered. For purposes of control and intervention, such a requirement is often impractical. It has also been criticized as failing to match our intuitions in causal reasoning: we would like to uphold the claim that smoking causes lung cancer, even if there is a subpopulation with a rare gene, where smoking actually decreases the probability of lung cancer (Dupré, 1984).

The **interventionist account of causation** (Spirtes et al., 2000; Pearl, 2000; Woodward, 2003) provides an alternative to a purely probabilistic model of causation. It supplements the probabilistic account with counterfactual thinking. On the interventionist account, causal reasoning is always relative to the choice of a causal model M: a directed acyclical graph (DAG) G, consisting of a set of vertices (=variables) and directed edges, and a probability distribution over the variables in G. Figure 1 is a

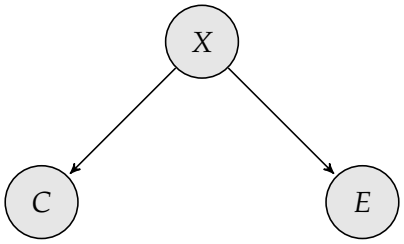

Figure 1: A typical common cause (conjunctive fork) structure. Conditional on X, the variables C and E are independent. Moreover, an intervention on C would disrupt the causal arrow from X to C and not have any effect on E.

simple example of a DAG. The edges represent how an intervention on a variable transfers to other variables; conversely, lack of a direct connection between variables codifies a conditional independence. DAGs can also codify Pearl's "causal assumptions" underneath our causal reasoning, when the arrows are interpreted causally. For instance, each variable is assumed to be independent of its non-descendants, given its direct causes (=its parents)—this is the famous Causal Markov Condition. Causally interpreted DAGs are called **causal Bayes nets**.

On the interventionist account, C is a cause of E if and only if an **intervention** on C causes a change of value in E, or changes the probability that E takes a certain value (Woodward, 2012). Note that the interventionist account conceives causation as a relationship between variables, whereas the probabilistic account describes relationships between events, or instantiations of such variables.[1]

But what is an intervention in the first place? An ideal intervention forces a variable C to take a certain value while breaking the influence of other variables. In a DAG, this amounts to breaking all arrows that lead into that variable. 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

[1]In this paper, I italicize all propositional variables; their instantiations are printed in roman script, following Bovens and Hartmann (2003).

light depends in turn 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 preconfigured system settings.

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. When one intervenes on C, the causal arrow leading from X to C is broken and no effect on E occurs. See Figure 1. While the probabilistic account compares $p(\mathrm{E}|\mathrm{C})$ and $p(\mathrm{E}|\neg\mathrm{C})$, the interventionist account focuses on probability of the effect conditional on an intervention on the cause (see also Meek and Glymour, 1994).

By explicating causal strength relative to causal models represented by causal Bayes nets, this paper combines both perspectives. Causal Bayes nets make it easy to spot which interventions affect which variables, and they have proved to be extremely useful for causal inference in science (e.g., Pearl, 2000; Spirtes et al., 2000). On top of this, they resemble other tools for causal inference, such as neural networks or connectionist expert systems. All in all, explicating causal strength relative to causal Bayes nets looks like a sound methodological assumption to make (see also Korb et al., 2011).

However, I also take over some elements from the probabilistic account of causation. First, I define causal strength as a relationship between propositions rather than variables, which is inspired by the question "how much does the cause raise the probability of the effect?". Second, statistical relevance plays a major role in the explication of causal strength measures. The next section fleshes out the details of my hybrid account.

2.2 Explicating the Determinants of Causal Strength

As stated above, I understand causal strength as a relationship between propositions: claims about the values of variables in a causal model, such as $X = x$, $Y = y$, and so

on. To keep notation simple, I focus on binary propositional variables, but nothing prevents us from generalizing the results to propositions about multicategorical or metric (e.g., real-valued) variables. I explain implications for more complex cases of causal inference at the end of the paper.

Let $\mathcal{L}$ be a propositional language with elements C, E $\in \mathcal{L}$. The causal strength between C and E is defined relative to a causal model $M = (\mathrm{G}, p(\cdot)) \in \mathcal{M}$: a directed acyclical graph G which includes the propositional variables $C \in \{\mathrm{C}, \neg\mathrm{C}\}$ and $E \in \{\mathrm{E}, \neg\mathrm{E}\}$, and a probability distribution $p(\cdot)$ over the variables in G. The set of causal models over variables from $\mathcal{L}$ is denoted by $\mathcal{M}$. 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).

After these preliminaries, we are now ready for describing the determinants of causal strength.

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

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

In particular, the causal strength measure $\eta(\mathrm{C}, \mathrm{E})$ can be represented by a continuous function $f : [0, 1]^3 \to \mathbb{R}$ such that

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

Here, $do(\mathrm{C})$ and $do(\neg\mathrm{C})$ are convenient notational shortcuts for setting the binary variable C to its values C and ¬C. For reasons of convenience, reference to a causal model is dropped in the notation of η. Also, in the adequacy conditions to follow, we will always assume that M contains the variables C and E when we quantify over C,

E and M.

Formality fixes the formal structure and the determinants of a causal strength measure $\eta(\mathrm{C}, \mathrm{E})$. In particular, it can be expressed as a function of the base rate of the cause and the probability of E under the relevant interventions on the cause: $p(\mathrm{E}|do(\mathrm{C}))$ and $p(\mathrm{E}|do(\neg\mathrm{C}))$. This takes up the basic idea behind probabilistic relevance accounts of causation, which is essentially contrastive. More precisely, we express causal strength as a function of $p(\mathrm{C})$, $p(\mathrm{E}|do(\mathrm{C}))$ and $p(\mathrm{E}|do(\neg\mathrm{C}))$, given a particular background context. Many of our measures will also maintain the intuition that positive causes need to raise the probability of their effects: C causes E if and only if $p(\mathrm{E}|do(\mathrm{C})) > p(\mathrm{E}|do(\neg\mathrm{C}))$, and C prevents E if and only if $p(\mathrm{E}|do(\mathrm{C})) < p(\mathrm{E}|do(\neg\mathrm{C}))$. The only difference to the venerable probabilistic account consists in the fact that we use the interventionist calculus instead of simple conditional probabilities.

Formality does not distinguish between causation at the population level (=event-types) and causation at the individual level (=singular events). The probabilistic-interventionist calculus is suited for describing causal strength at both levels. This distinction does not cover actual causation, or "cause in fact" (Halpern and Pearl, 2005a,b). In those cases, one already knows that both the cause and the effect occurred: instead of estimating the predictive value of the cause, one has to attribute the effect to one of the causes, and to compare the contribution of different causes. Most results below are not meant to apply to actual causation.

Notably, Formality 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, 2015). While this implies that our model does not capture all factors that may influence 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 estimate causal strength from observations and the results of interventions alone.

Formality is also blind to mediator variables—variables on a path between C and

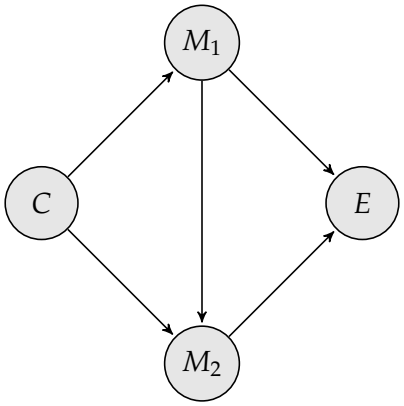

Figure 2: Mediators on the paths linking cause C and effect E.

E—and the presence of multiple paths leading from *C* to *E* (see Figure 2). Also this choice is conscious. The reason is that mediators are often not directly measurable. When we administer a medical drug (C) to cure migraine (E), there are numerous mediators in an appropriate causal model that includes C and E. However, the medical practitioner, who has to choose between different drugs, is mainly interested in the overall effect that C has on E (how often does migraine go away?), not in the details of causal transmission within the human body. Therefore I keep the model simple and amalgamate the effects that C may have on E via different paths into one number—the average or total effect of C on E (e.g., Dupré, 1984; Eells, 1991).

Sometimes, measuring path-specific effects is vital for policy-making and causal attribution. My decision to focus on total effect does not rule out a path-specific perspective, quite to the contrary. Measures of path-specific effect supervene on elementary measure of causal strength (e.g., Pearl, 2001). Only when such a measure is chosen, one can get the calculations off the ground and meaningfully separate direct from indirect effect. By investigating the properties of these elementary measures, this paper grounds any template for measuring path-specific causal strength.

Table 2, which follows Fitelson and Hitchcock (2011), translates various probabilistic measures of causal strength to the causal Bayes nets framework. By investigating their formal properties, the following section creates a basis for comparing, discussing

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

Table 2: Some prominent measures of causal strength. I follow the labels of Fitelson and Hitchcock (2011).

and appraising these measures. In the end, I will also defend my personal preferences and draw some tentative conclusions regarding the question of whether we should work with a single causal strength measure, or a plurality of measures. The dialectics of the argument go either way: on the one hand, I use plausible adequacy conditions to back a specific measure, on the other hand, I use implausible adequacy conditions to argue against competitors.

3 The Representation Theorems

3.1 General Adequacy Conditions

The previous section has determined the mathematical structure and determinants of measures of causal strength. The adequacy conditions in this subsection deal with general properties of the measures, in particular with their ranking of different cause/effect pairs.

We start with comparing two putative causes of an effect E. Suppose we ask what

is a stronger cause of car accidents (E): drunk driving (C_1) or bad weather conditions (C_2)? One may answer that C_1 is a stronger cause of E than C_2 if and only if C_1 makes E more expected than C_2. In other words, a cause of an effect is stronger than another cause if it has a higher likelihood of producing the effect. Consequently, we judge drunk driving to be a stronger cause of car accidents than bad weather. The following condition describes that way of reasoning.

Effect Production For propositions C_1, C_2, and $E \in \mathcal{L}$ and a causal model $M \in \mathcal{M}$,

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

It may be objected that Effect Production neglects the contrastive nature of measures of causal strength. They should measure the degree of causal dependence on C as opposed to $\neg$C, that is, the difference that an intervention on C makes for E. Measures that satisfy Effect Production, however, are not sensitive to what happens if $\neg C_1$ or $\neg C_2$ is the case. On the other hand, production rather than dependence is an important part of many theories of causation (e.g., Dowe, 2000; Hall, 2004) and measures of causal strength that focus on the probability of producing E should not be dismissed from the start.

Inspired by the above concerns, the next adequacy condition focuses on the difference that two competing causes make for the target effect:

Difference-Making For propositions C_1, C_2, and $E \in \mathcal{L}$ and a causal model $M \in \mathcal{M}$,

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

if and only if for a continuous function $g : [0,1]^2 \to \mathbb{R}$ which is monotonically increasing in the first and monotonically decreasing in the second argument, without being constant in either argument:

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

This condition adopts a contrastive perspective on causal strength: we only look at the difference which two interventions make for the effect.[2] The monotonicity constraint on g expresses the intuitive condition that the more likely a cause C is to bring about an effect E, the more sizeable its causal effect, all other things being equal. Vice versa for $\neg$C. Such ways of causal reasoning are particularly relevant for RCTs and other contexts where treatment (C) and control ($\neg$C) groups are compared with respect to their causal effect, e.g., for comparing the strength of two medical drugs.

We now proceed from the determinants to the scaling properties of causal strength measures. I follow Fitelson and Hitchcock (2011) in explicating the degree to which C prevents E as the degree to which C causes $\neg$E, that is, the absence of E. To measure causation and prevention on the same scale, we now demand that the causal strength of C for E equal the causal strength of C for $\neg$E, in terms of its absolute values. The idea is that the negative causal effect of C on E (=the degree of causal prevention) is just the flip side of the positive effect of C on $\neg$E; we are dealing with one and the same propositional variable. This leads us to Fitelson and Hitchcock's

Causation-Prevention Symmetry (CPS) For propositions C, E $\in \mathcal{L}$ and a causal model $M \in \mathcal{M}$,

$$-\eta(\mathrm{C},\mathrm{E}) = \eta(\mathrm{C},\neg\mathrm{E})$$

Evidently, only measures of causal strength which take both positive and negative values can satisfy CPS. Positive causal strength indicates positive causation, negative causal strength indicates prevention. As a consequence, $\eta(\mathrm{C},\mathrm{E}) = 0$ denotes neutral causal strength.[3] A purely ordinal version of CPS is the following, strictly weaker condition:

Weak Causation-Prevention Symmetry (WCPS) For propositions C, E_1, and $\mathrm{E}_2 \in \mathcal{L}$

[2] Base rates of causes may still matter for the effect. For instance, if C_1 and C_2 are two independent causes of E, then $p(\mathrm{E}|do(\mathrm{C}_2)) = p(\mathrm{E}|do(\mathrm{C}_2),\mathrm{C}_1)p(\mathrm{C}_1) + p(\mathrm{E}|do(\mathrm{C}_2),\neg\mathrm{C}_1)p(\neg\mathrm{C}_1)$ takes into account the base rate of the competing cause C_1.

[3] This should not be conflated with causal irrelevance. A cause can be relevant for an effect, and yet, the overall effect can be zero, e.g., when contributions via different paths cancel out. This is different from a case where there is no causal connection between C and E.

and a causal model $M \in \mathcal{M}$, let E_1 and E_2 be screened off by a common cause C (i.e., $E_1 \perp\!\!\!\perp E_2$ given C and $\neg$C). Then

$$\eta(C, E_1) > \eta(C, E_2) \qquad \text{if and only if} \qquad \eta(C, \neg E_1) < \eta(C, \neg E_2)$$

This condition demands that if C is a stronger cause of E_1 than of E_2, then C prevents $\neg E_1$ more than $\neg E_2$. The common cause condition has been added to avoid interactions between E_1 and E_2 which might confound our intuitions.

I will often rescale candidate measures into a form that satisfies CPS. Moreover, to avoid that the discussion depends too much on specific scaling properties, I will usually not compare individual measures, but equivalence classes of measures, characterized by the causal strength rankings they impose. The key concept is **ordinal equivalence**. Two measures η and η' are ordinally equivalent if and only if for all propositions C, C′, E, E′ $\in \mathcal{L}$ and causal models $M \in \mathcal{M}$,

$$\eta(C, E) > \eta(C', E') \qquad \text{if and only if} \qquad \eta'(C, E) > \eta'(C', E')$$

relative to M. Two ordinally equivalent measures may use different scales, but they produce the same causal strength rankings and share most philosophically interesting properties. The following subsections provide representation theorems for classes of ordinally equivalent measures.

3.2 Causal Production and the Pearl Measure

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

In this section, I provide an axiomatic characterization of Pearl's measure of causal effect $\eta(\mathrm{C},\mathrm{E}) = p(\mathrm{E}|do(\mathrm{C}))$ (Pearl, 2000). To this end, I introduce another adequacy condition. Consider Table 3. Three teams in the Italian *Seria A*—AS Roma, FC Internazionale ("Inter"), and Juventus ("Juve")—are still 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. (Juve misses four and five points on both teams and cannot surpass them any more.) In that case, we might say that C causes E_1 and E_2 to an equal degree: given C, we cannot distinguish between both propositions. This intuition is stated in the following adequacy condition:

Conditional Equivalence For C, E_1 and $\mathrm{E}_2 \in \mathcal{L}$ and a causal model $M \in \mathcal{M}$, assume that E_1 and E_2 are logically equivalent given C (i.e., $\mathrm{E}_1 \wedge \mathrm{C} \models \mathrm{E}_2$ and $\mathrm{E}_2 \wedge \mathrm{C} \models \mathrm{E}_1$). Then $\eta(\mathrm{C},\mathrm{E}_1) = \eta(\mathrm{C},\mathrm{E}_2)$.

Taking this condition together with Formality and Effect Production, we can prove the following theorem:

Theorem 1 (Representation Theorem for η_p) *All measures of causal strength that satisfy Formality, Effect Production and Conditional Equivalence are ordinally equivalent to*

$$\eta_p(\mathrm{C},\mathrm{E}) = p(\mathrm{E}|do(\mathrm{C}))$$

Pearl (2000, 70) calls $\eta_p(\mathrm{C},\mathrm{E}) = p(\mathrm{E}|do(\mathrm{C}))$ the "causal effect" of C on E. It is easily generalizable to the case to the causal strength of an intervention for a metric variable E (rather than a proposition), where causal strength corresponds to the average value of E under an intervention on C: $\eta_P(\mathrm{C},E) = \mathbb{E}[E|do(\mathrm{C})]$ (ibid.).

Pearl's measure also fits well with cases of causal attribution. Kaiserman (2016a,b) suggests that when several causes contribute to an effect, degree of causal contribution and partial liability in legal reasoning should be proportional to η_p. The driver's

drunkenness (C_1) has a much higher tendency to produce an accident (E) than the bad weather conditions (C_2); our judgments of causal strength and causal contribution should mirror this fact. Kaiserman also applies a normalized version of η_p to voting bodies in order to measure the causal contributions of different parties to a decision (cf. Braham and van Hees, 2009). Note, however, that η_p violates Difference-Making, and it does not distinguish between (positive) causation, causal prevention, and causal neutrality.

3.3 Separability of Effects and the Probability-Raise Measure

The probabilistic causation literature identifies causes by their raising of the probability of an effect. C raises the probability of E just in case

$$p(\mathrm{E}|\mathrm{C}) > p(\mathrm{E}). \tag{2}$$

While some authors, such as Skyrms (1980) and Eells (1991), prefer to identify probability-raising in a strictly contrastive way (i.e., $p(\mathrm{E}|\mathrm{C}) > p(\mathrm{E}|\neg\mathrm{C})$), Equation (2) is probably the most venerable and intuitive expression of the idea of probability-raising (Reichenbach, 1956; Suppes, 1970; Cartwright, 1979). The corresponding measure of causal strength is, in the interventionist notation,

$$\eta_{pr}(\mathrm{C},\mathrm{E}) = p(\mathrm{E}|do(\mathrm{C})) - p(\mathrm{E}). \tag{3}$$

While this measure has never been defended systematically (though see the cursory reference to η_{pr} in Pearl, 2011, 717), it is a natural candidate for measuring causal strength: it quantifies the extent to which an intervention on C raises the base rate of E. Indeed, η_{pr} can be characterized in a simple and straightforward way, using the following property:

Separability of Effects For $\mathrm{C} \in \mathcal{L}$, two mutually exclusive propositions $\mathrm{E}_1, \mathrm{E}_2 \in \mathcal{L}$ and a causal model $M \in \mathcal{M}$, there is a function $f : \mathbb{R}^2 \to \mathbb{R}$, monotonically

increasing in each argument, such that

$$\eta(\mathrm{C}, \mathrm{E}_1 \vee \mathrm{E}_2) = f(\eta(\mathrm{C}, \mathrm{E}_1), \eta(\mathrm{C}, \mathrm{E}_2))$$

with

$$\begin{aligned}
&\eta(\mathrm{C}, \mathrm{E}_1 \vee \mathrm{E}_2) > \eta(\mathrm{C}, \mathrm{E}_1) && \text{if and only if} && p(\mathrm{E}_2|do(\mathrm{C})) > p(\mathrm{E}_2) \\
&\eta(\mathrm{C}, \mathrm{E}_1 \vee \mathrm{E}_2) = \eta(\mathrm{C}, \mathrm{E}_1) && \text{if and only if} && p(\mathrm{E}_2|do(\mathrm{C})) = p(\mathrm{E}_2) \\
&\eta(\mathrm{C}, \mathrm{E}_1 \vee \mathrm{E}_2) < \eta(\mathrm{C}, \mathrm{E}_1) && \text{if and only if} && p(\mathrm{E}_2|do(\mathrm{C})) < p(\mathrm{E}_2)
\end{aligned}$$

and vice versa for E_1 and E_2 reversed.

Separability of Effects is an eminently intuitive and practical principle. For example, if E_1 and E_2 state that the value of a variable X is in the (disjoint) intervals I_1 and I_2, respectively, we can use Separability of Effects to calculate the causal strength of C on $X \in I_1 \cup I_2$ from the causal strength of C for E_1 and E_2. Moreover, it states that the causal strength for the disjunction is greater than the causal strength for the individual effects just in case C is a (positive) cause of either effect.

With the help of Separability of Effects, we can characterize η_{pr} up to ordinal equivalence:

Theorem 2 (Representation Theorem for η_{pr}) *All measures of causal strength that satisfy Formality, Effect Production and Separability of Effects are ordinally equivalent to*

$$\eta_{pr}(\mathrm{C}, \mathrm{E}) = p(\mathrm{E}|do(\mathrm{C})) - p(\mathrm{E})$$

As an additional benefit, η_{pr} satisfies the Causation-Prevention Symmetry CPS.

We now proceed to representation theorems for measures which violate Effect Production and satisfy Difference-Making instead. Given Formality and Difference-Making, each of the adequacy conditions discussed below is sufficient to single out a measure of causal strength up to ordinal equivalence. In this specific sense, the

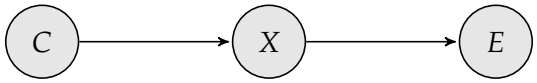

Figure 3: A DAG representing causation along a single path.

following conditions are pairwise inconsistent with each other.

3.4 Multiplicativity on Single Paths and the Difference Measure

How should causal strength combine on a single path, e.g., in the graph in Figure 3? If causal strength is conceived of as the intensity of a physical process linking cause and effect, overall causal strength should be a function of the causal strength between the individual links. But which function $g : \mathbb{R}^2 \to \mathbb{R}$ should be chosen such that $\eta(\mathrm{C}, \mathrm{E}) = g(\eta(\mathrm{C}, \mathrm{X}), \eta(\mathrm{X}, \mathrm{E}))$?

A couple of requirements suggest themselves. First of all, g should be symmetric: the order of mediators in a chain does not matter. Whether a weak link precedes a strong link, or vice versa, should not matter for overall causal strength. Second, it seems that the overall causal strength cannot be stronger than the weakest link in the chain: If C and X are almost independent, it does not matter how strongly X and E are correlated: the causal strength will be still 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., $\eta(\mathrm{C}, \mathrm{X}) = 1$), then the strength of the entire chain will just be the strength of the rest of the chain. Perfect connections between two nodes neither raise nor attenuate overall causal strength (see also Good, 1961a, 311–312).

A very simple function that satisfies all these requirements is multiplication. Thus, I propose the following principle:

Multiplicativity on Single Paths If the variables C and E are connected via a single path with intermediate node X, then $\eta(\mathrm{C}, \mathrm{E}) = \eta(\mathrm{C}, \mathrm{X}) \cdot \eta(\mathrm{X}, \mathrm{E})$.

As a corollary, we obtain that for a causal chain with multiple mediators, e.g.,

$C \to X_1 \to \ldots \to X_n \to E$,

$$\eta(\mathrm{C}, \mathrm{E}) = \eta(\mathrm{C}, \mathrm{X}_1) \cdot \eta(\mathrm{X}_1, \mathrm{X}_2) \cdot \ldots \cdot \eta(\mathrm{X}_{\mathrm{n}-1}, \mathrm{X}_{\mathrm{n}}) \cdot \eta(\mathrm{X}_{\mathrm{n}}, \mathrm{E})$$

In many cases, causal reasoning follows that principle. For instance, taking a medical drug (C) may lower cholesterol levels substantially (X), but the effect of the drug on mortality (E) via its impact on cholesterol may still be minuscule—just because cholesterol is only one of many causes of mortality. Incidentally, Multiplicativity on Single Paths assumes transitivity of causation, which is warranted for the binary variable models considered in this paper (Korb et al., 2011; Halpern, 2016).

It is now possible to characterize all measures that satisfy Multiplicativity on Single Paths, in addition to Formality and Difference-Making:

Theorem 3 (Representation Theorem for η_d) *All measures of causal strength that satisfy Formality, Difference-Making and Multiplicativity on Single Paths are ordinally equivalent to*

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

This is a simple and intuitive quantity that measures causal strength by comparing the effect that different interventions on C have on E. Indeed, η_d is straightforwardly applicable in statistical inference. For example, in clinical trials and epidemiological studies, $\eta_d(\mathrm{C}, \mathrm{E})$ reduces to Absolute Risk Reduction, or ARR.

Holland (1986, 947) calls η_d the "average causal effect" of C on E—a label that is motivated by the fact that η_d aggregates the strength of different causal links. Pearl (2001) uses η_d as the basis for developing a theory of direct and indirect causal effects. Neither of them justifies the choice of η_d on foundational grounds—a gap that is closed by Theorem 3.

Note that η_d also satisfies CPS and Separability of Effects—in fact, it is the only measure based on Difference-Making that satisfies this property. Moreover, η_d can be

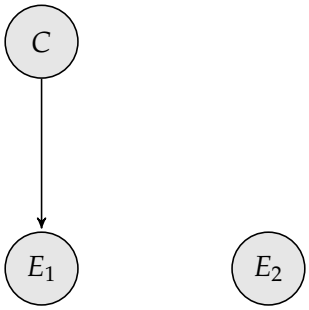

Figure 4: An effect E_2 which is irrelevant regarding the causal relation between C and E_1.

rewritten as

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

Modulo subtraction of a constant, $\eta_d(\mathrm{C},\mathrm{E})$ 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: $p(\neg\mathrm{E}|do(\neg\mathrm{C}))$ indicates to what extent C is *necessary* for producing E (what happens if an intervention brings about $\neg$C?), and $p(\mathrm{E}|do(\mathrm{C}))$ indicates to what extent the presence of C is *sufficient* for producing E (what happens if an intervention brings about C?). $\eta_d(\mathrm{C},\mathrm{E})$ intuitively combines these two plausible ways of thinking about causal strength.

3.5 No Dilution for Irrelevant Effects and Probability Ratio Measures

How strongly does C cause the conjunction of two effects—$\mathrm{E}_1 \wedge \mathrm{E}_2$—when C affects only one of them positively, and the other effect (say, E_2) is independent of C and of E_1? In such circumstances, we may call E_2 an "irrelevant effect". This situation is represented visually in the DAG of Figure 4.

There are two basic intuitions about what such effects mean for overall causal strength: either causal strength is diluted when passing from E_1 to $\mathrm{E}_1 \wedge \mathrm{E}_2$, or it is not. Dilution means that adding E_2 to E_1 diminishes causal strength, that is, $\eta(\mathrm{C},\mathrm{E}_1 \wedge \mathrm{E}_2) <$

$\eta(\mathrm{C}, \mathrm{E}_1)$. Conversely, a measure is non-diluting if and only if in these circumstances, $\eta(\mathrm{C}, \mathrm{E}_1 \wedge \mathrm{E}_2) = \eta(\mathrm{C}, \mathrm{E}_1)$. This amounts to the following principle:

No Dilution for Irrelevant Effects For propositions $\mathrm{C}, \mathrm{E}_1, \mathrm{E}_2 \in \mathcal{L}$ and a causal model $M \in \mathcal{M}$, let $E_2 \perp\!\!\!\perp C$, and $E_2 \perp\!\!\!\perp E_1$ conditional on C. Then $\eta(\mathrm{C}, \mathrm{E}_1 \wedge \mathrm{E}_2) = \eta(\mathrm{C}, \mathrm{E}_1)$.

Non-diluting measures of causal strength that satisfy Difference-Making can be neatly characterized.[4] In fact, they are all ordinally equivalent to Lewis' probability ratio measure (Lewis, 1986), as the following theorem demonstrates.

Theorem 4 (Representation Theorem for η_r) *All measures of causal strength that satisfy Formality, Difference-Making, and No Dilution for Irrelevant Effects are ordinally equivalent to*

$$\eta_r(\mathrm{C}, \mathrm{E}) = \frac{p(\mathrm{E}|do(\mathrm{C}))}{p(\mathrm{E}|do(\neg\mathrm{C}))}$$

and its rescaling to the $[-1; 1]$ *range*

$$\eta_{r'}(\mathrm{C}, \mathrm{E}) = \frac{p(\mathrm{E}|do(\mathrm{C})) - p(\mathrm{E}|do(\neg\mathrm{C}))}{p(\mathrm{E}|do(\mathrm{C})) + p(\mathrm{E}|do(\neg\mathrm{C}))}.$$

To some extent, this result can be interpreted as a *reductio ad absurdum* of the probability ratio measure family. After all, given the lack of a causal connection between C and E_2, it is plausible that C causes $\mathrm{E}_1 \wedge \mathrm{E}_2$ to a smaller degree than E_1. Rain in New York on November 26, 2016 (C) affects umbrellas sales in that city (E_1), but it does not affect whether FC Barcelona will win their next Champions League match (E_2). Therefore, the causal effect of rain on umbrella sales should be stronger than the causal effect on umbrella sales in conjunction with Barcelona winning their next match. This is bad news for η_r and $\eta_{r'}$.

A way around this problem consists in restricting No Dilution for Irrelevant Effects to prevention rather than (positive) causation. According to the intuition that underlies CPS and WCPS, if C is a strong preventive cause of E_1, then it is a strong positive cause

[4]Incidentally, the premises of the No Dilution condition are compatible with a prima facie correlation between E_1 and E_2. However, this correlation vanishes as soon as we control for different levels of C.

of $\neg \mathrm{E}_1$. Changing the prevented effect from E_1 to $\mathrm{E}_1 \wedge \mathrm{E}_2$ then amounts to changing the positively caused effect from $\neg \mathrm{E}_1$ to $\neg(\mathrm{E}_1 \wedge \mathrm{E}_2) = \neg \mathrm{E}_1 \vee \neg \mathrm{E}_2$. In other words, we make the (positive) effect less specific. It is not clear why this relaxation should attenuate causal strength, and for this reason, degree of prevention should not be diluted by adding irrelevant effects.[5] We obtain the following adequacy condition:

No Dilution for Irrelevant Effects (Prevention) For propositions $\mathrm{C}, \mathrm{E}_1, \mathrm{E}_2 \in \mathcal{L}$ and a causal model $M \in \mathcal{M}$ let $E_2 \perp\!\!\!\perp C$, $E_2 \perp\!\!\!\perp E_1$ conditional on C, and let C be a preventive cause of E_1. Then $\eta(\mathrm{C}, \mathrm{E}_1 \wedge \mathrm{E}_2) = \eta(\mathrm{C}, \mathrm{E}_1)$.

I have motivated No Dilution for Irrelevant Effects (Prevention) by means of causation-prevention symmetries; so it is natural that both conditions should go together. Indeed, they figure in one and the same representation theorem.

Theorem 5 (Representation Theorem for η_{cg}) *All measures of causal strength that satisfy Formality, Difference-Making, No Dilution for Irrelevant Effects (Prevention) and Weak Causation-Prevention Symmetry are ordinally equivalent to*

$$\eta_{cg}(\mathrm{C}, \mathrm{E}) = \begin{cases} \frac{p(\mathrm{E}|do(\mathrm{C})) - p(\mathrm{E}|do(\neg \mathrm{C}))}{1 - p(\mathrm{E}|do(\neg \mathrm{C}))} & \text{if } C \text{ is a positive cause of } E \\ \\ \frac{p(\mathrm{E}|do(\mathrm{C})) - p(\mathrm{E}|do(\neg \mathrm{C}))}{p(\mathrm{E}|do(\neg \mathrm{C}))} & \text{if } C \text{ is a preventive cause of } E \end{cases}$$

For the case of positive causation, this measure agrees with two prominent proposals from the literature. The psychologist Patricia Cheng (1997) derived

$$\eta_c(\mathrm{C}, \mathrm{E}) := \frac{p(\mathrm{E}|do(\mathrm{C})) - p(\mathrm{E}|do(\neg \mathrm{C}))}{1 - p(\mathrm{E}|do(\neg \mathrm{C}))} \tag{4}$$

(that is, η_{cg} without the above case distinction) from theoretical considerations about how agents perform causal induction and called it the "causal power" of C on E.

[5] I owe this line of argument to Sander Beckers. Moreover, there is an analogy to Vincenzo Crupi and Katya Tentori's discussion of probabilistic measures of explanatory power: they argue that explanatory power should be diluted for adding irrelevant explananda to a good explanation, but they plea for non-dilution in the case of explanatory failure (Crupi and Tentori, 2012, 369–371)—see also Cohen (2016).

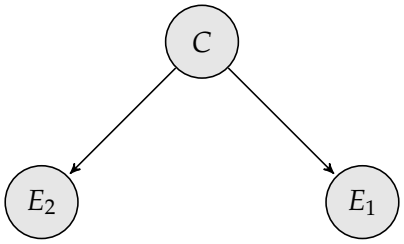

Figure 5: A typical common cause structure where C screens off the two effects E_1 and E_2.

Cheng's measure is in turn ordinally equivalent to the measure

$$\eta_g(\mathrm{C},\mathrm{E}) = \frac{p(\neg \mathrm{E}|do(\neg \mathrm{C}))}{p(\neg \mathrm{E}|do(\mathrm{C}))} = \frac{1 - p(\mathrm{E}|do(\neg \mathrm{C}))}{1 - p(\mathrm{E}|do(\mathrm{C}))}$$

that the statistician and philosopher of science I.J. Good (1961a,b) derived from a complex set of adequacy conditions. This ordinal equivalence, noted first by Fitelson and Hitchcock (2011), is evident from the equation below.

$$\eta_c(\mathrm{C},\mathrm{E}) = \frac{-p(\neg \mathrm{E}|do(\mathrm{C})) + p(\neg \mathrm{E}|do(\neg \mathrm{C}))}{p(\neg \mathrm{E}|do(\neg \mathrm{C}))} = -\frac{1}{\eta_g(\mathrm{C},\mathrm{E})} + 1$$

The two previous theorems elucidate that η_r and η_{cg} are based on the same principle: the No Dilution for Irrelevant Effects property. The crucial question which separates the two classes of measures is whether this property should hold across the board or just for preventive causes.

3.6 Conjunctive Closure and the Logarithmic Ratio Measure

Consider a variable C that affects a set of other variables E_1, E_2, ..., which would be unrelated to each other, were it not for their common cause C. In this scenario, represented visually in Figure 5, one could ask whether the and the individual effects (E_1, E_2) constrain the strength of the causal link between C and the conjunction of these variables. In other words, we ask how $\eta(\mathrm{C}, \mathrm{E}_1 \wedge \mathrm{E}_2)$ depends on $\eta(\mathrm{C}, \mathrm{E}_1)$ and $\eta(\mathrm{C}, \mathrm{E}_2)$, and under which circumstances the former is a function of the latter.

Imagine, for example, that a medical drug has two side effects—diarrhea and sore throat—which are independent of each other. Both side effects are caused with the same strength t. It is then natural to say that the overall side effect of the medical drug is also equal to t since there is no interaction between both effects. Apart from the intuitive plausibility, this principle facilitates practical calculations because we can often infer the strength of a cause C for an aggregate effect from the strength of C for the individual effects. This is similar to the principle of Separability of Effects, but now for a conjunction rather than a disjunction of effects. Formally:

Conjunctive Closure For propositions C, E_1 and $\mathrm{E}_2 \in \mathcal{L}$, and a causal model $M \in \mathcal{M}$ with $E_1 \perp\!\!\!\perp E_2$, the following implication holds:

$$\eta(\mathrm{C}, \mathrm{E}_1) = \eta(\mathrm{C}, \mathrm{E}_2) = t \qquad \Rightarrow \quad \eta(\mathrm{C}, \mathrm{E}_1 \wedge \mathrm{E}_2) = t \tag{5}$$

Incidentally, this line of reasoning resembles the Conjunction Principle in epistemology, which states that justification and/or knowledge are closed under logical conjunction. Shogenji (2012) transfers this principle to the quantitative dimension and to probabilistic measures of justification. When (i) the degree of justification of H_1 and H_2 given E is both equal to t and (ii) H_1 and H_2 are probabilistically independent (unconditionally and conditionally on E), then the degree of justification of $\mathrm{H}_1 \wedge \mathrm{H}_2$ should also be equal to t. Put differently, justification is not diluted under the conjunction of independent propositions.[6] Shogenji (2012) calls this the Special Conjunction Principle, and Conjunctive Closure demands a similar property for measures of causal strength.

It is possible to characterize measures which satisfy Conjunctive Closure up to ordinal equivalence (see also Atkinson, 2012):

Theorem 6 (Representation Theorem for η_{cc}) *All measures of causal strength that satisfy*

[6]Without the independence assumption, this principle would immediately collapse due to lottery paradox-type counterexamples.

Formality, Difference-Making and Conjunctive Closure are ordinally equivalent to

$$\eta_{cc}(\mathrm{C},\mathrm{E}) = \frac{\log p(\mathrm{E}|do(\mathrm{C}))}{\log p(\mathrm{E}|do(\neg\mathrm{C}))}$$

Although this measure has not yet been proposed in the literature, it is based on a reasonable principle, and therefore a serious candidate for measuring causal strength.

This result concludes our axiomatic characterization of measures of causal strength. We now proceed to a practical application and a comparison of the candidate measures.

4 Application: Quantifying Causal Effect in Medicine

A natural scientific application of probabilistic measures of causal strength consists in quantifying the size of an effect that an intervention on one variable has on another. Randomized Controlled Trials (RCTs) in clinical medicine, or cohort studies in epistemology, are particularly salient applications. As we have seen in the introduction, the medical literature uses different measures to quantify causal strength. They can be translated into our framework by writing observed relative frequencies as conditional probabilities under an intervention. In particular, the measures we discussed in the introduction—relative risk, relative risk reduction, absolute risk reduction—can be given the following reading:

$$\mathrm{RR} = \frac{p(\mathrm{E}|do(\mathrm{C}))}{p(\mathrm{E}|do(\neg\mathrm{C}))} \qquad \text{(Relative Risk)}$$

$$\mathrm{ARR} = p(\mathrm{E}|do(\mathrm{C})) - p(\mathrm{E}|do(\neg\mathrm{C})) \qquad \text{(Absolute Risk Reduction)}$$

$$\mathrm{RRR} = \frac{p(\mathrm{E}|do(\mathrm{C})) - p(\mathrm{E}|do(\neg\mathrm{C}))}{p(\mathrm{E}|do(\neg\mathrm{C}))} \qquad \text{(Relative Risk Reduction)}$$

It is not difficult to relate these measures to measures of causal strength. For

example, RR is just the familiar probability ratio measure η_r, whereas ARR turns out to be the difference measure η_d. RRR = RR - 1 is ordinally equivalent to η_r, too.

Apart from the mathematical characterization, also normative arguments in favor or against causal strength measures carry over to these effect size measures. Since the probability ratio measure η_r satisfies No Dilution for Irrelevant Effects, so do RR and RRR. The value of those measures does not change when irrelevant propositions are added to the effect. This can have extremely undesirable consequences, also from a practical point of view. The causal effect of a painstiller on relieving headache is, according to η_r, as big as the causal effect of that drug on relieving headache *and* a completely unrelated phenomenon, e.g., lowering cholesterol levels. Such properties are grossly misleading and may have devastating practical consequences. A drug which is highly effective for curing one of two unrelated symptoms will be mistaken for curing both symptoms at once, even if the second symptom is causally disconnected from the drug. Hence, our formal results undermine the use of those measures in clinical practice. These conclusions exceed the realm of purely causal inference and also concern observational studies, e.g., measuring the risk that certain dietary habits pose for two completely different kinds of cancer.

On the other hand, the defining properties of η_d, in particular Multiplicativity on Single Paths, suit clinical practice well. We have already seen the example that taking a medical drug (C) may lower cholesterol levels substantially (X), but the effect of the drug on mortality (E) via its impact on cholesterol is minuscule. Combining causal strength along a single path with the formula $\eta_d(\mathrm{C},\mathrm{E}) = \eta_d(\mathrm{C},\mathrm{X}) \cdot \eta_d(\mathrm{X},\mathrm{E})$ takes this into account: the overall causal strength will be very small if one of the links is tenuous.

Finally, our theoretical arguments nicely square with decision-theoretic and epistemic arguments for preferring absolute over relative risk measures in medicine, e.g., the neglect of base rates in relative risk measures (Stegenga, 2015). Pursuing this topic in detail deserves a separate paper, but one feature should be listed. Suppose the treatment group is given a new drug against influenza. The control group is given a placebo. Assume that we know the costs and benefits of fast recovery (E) as opposed

to delayed or no recovery ($\neg$E), as well as the costs of administering the treatment and the control (broadly construed, including harmful side effects). Then we can determine the optimal decision for future cases on the basis of the value of η_d in the observed sample. Here "optimal" is understood in the sense of maximizing expected utility. Among the measures investigated in this paper, η_d (respectively ARR) is the only measure that is sufficient for determining optimal decisions (Sprenger and Stegenga, 2016). Other measures require additional information, such as the base rate of recovery in the control group.

It should now be evident that probabilistic measures of causal strength have important applications in scientific inference and beyond. This analysis provides valuable guidance for evaluating those measures and choosing between them.

5 Discussion

This paper has provided axiomatic foundations for a probabilistic theory of causal strength. Synthesizing ideas from the probability-raising and the interventionist view of causation, I have proposed to evaluate causal strength relative to a causal Bayes net in which causes and effects are embedded. Causal strength is determined by the probability of an effect under interventions that bring about C and $\neg$C, $p(\mathrm{E}|do(\mathrm{C}))$ and $p(\mathrm{E}|do(\neg\mathrm{C}))$, and the base rate of the cause, $p(\mathrm{C})$. Both causes and effects have been understood as propositions rather than as events, objects or variables, and the evaluation of causal strength has taken place relative to a single background context.

After introducing the conceptual and mathematical framework, I have characterized various measures of causal strength in terms of various axioms and adequacy conditions. Such a characterization makes it possible to assess the merits of the different measures in the literature by means of assessing the plausibility of the adequacy conditions. Even if more than a single measure survives the theoretical scrutiny, one can still form informed preferences. While causal Bayes nets provide the framework for our analysis, the methods for characterizing the various measures, and for proving

the theorems, are partially transferred from various parts of formal epistemology, in particular Bayesian confirmation theory and Bayesian analyses of explanatory power.

Notably, the investigated measures fall into two major categories: those who satisfy Effect Production (=causes are ranked in terms of their tendency to produce the effect) and those who satisfy Difference-Making (=causal strength is an increasing function of $p(\mathrm{E}|do(\mathrm{C}))$ and a decreasing function of $p(\mathrm{E}|do(\neg\mathrm{C}))$). The Pearl measure $\eta_p(\mathrm{C},\mathrm{E}) = p(\mathrm{E}|do(\mathrm{C}))$ and the probability-raise measure $\eta_{pr}(\mathrm{C},\mathrm{E}) = p(\mathrm{E}|do(\mathrm{C})) - p(\mathrm{E})$ satisfy Effect Production, while the other measures violate it. The measures in that second group straightforwardly express the counterfactual dependence of E on C: how much does a change in the value of C affect the outcome E? Therefore they satisfy Difference-Making while η_p and η_{pr} violate that property.[7] These two measures are not based on the contrastive value that a cause has for an effect and may not be suitable for comparing the effect of different interventions on C. That said, they may be applicable when we are more interested in questions of attribution and liability than in the predictive value of a cause for an effect (e.g., Kaiserman, 2016a,b). In such cases, questions of causal production may be more relevant than questions of counterfactual dependence (see also Beckers, 2016).

To underscore the different angle of the discussed measures, consider a case of causal overdetermination (e.g., Lewis, 1973):

> An assassin puts poison into the king's wine glass (C). If the king does not drink the wine, a (reliable) backup assassin will shoot him. The king drinks the wine and dies (E).

Assuming the presence of the backup assassin as the relevant background context, the Pearl measure $\eta_p(\mathrm{C},\mathrm{E}) \approx 1$ judges the assassin's action as a strong cause of the king's death, even if the king's fate was sealed anyway. The measure $\eta_d(\mathrm{C},\mathrm{E})$, however, disagrees: since the assassin's action barely made a difference to the effect, $\eta_d(\mathrm{C},\mathrm{E}) \approx$

[7] η_{pr} is insensitive to $p(\mathrm{E}|do(\neg\mathrm{C}))$ if $p(\mathrm{C}) = 1$. Here we see an important difference between probabilistic measures of causal strength and probabilistic confirmation measures: in the latter field, that case would be classified as degenerate and unimportant, whereas there is no reason for doing so in causal reasoning. Causal strength assessments make sense for all base rates of the cause.

0. The two groups of measures will also diverge in cases where an action produces an effect, but by doing so, preempts an even stronger cause. Our probabilistic measures of causal strength thus neatly mirror different ways of understanding causal relevance.

In the light of these ambiguities, prospects may look bleak for the project of coming up with "the one true measure of causal strength". Also in the related field of confirmation theory, ambitions to establish a single probabilistic measure (e.g., Milne, 1996) did ultimately give way to a plurality of confirmation measures (e.g., Fitelson, 2001; Brössel, 2013; Crupi, 2013). But I contend that the situation is different for causal strength. Among the measures that are based on the Difference-Making principle, it is possible to form rational preferences that are more than personal taste. Take a look at the properties of the measures summarized in Table 4. It is notable that only two measures (η_d and η_{cg}) satisfy the Weak Causation-Prevention Symmetry, although this is an eminently sensible property. The same can be said about Multiplicativity on Single Paths, which is only satisfied by Eells' measure η_d. This is also the only measure based on Difference-Making which satisfies Separability of Effects, allowing for an easy decomposition of complex effects. On the other hand, Theorem 4 points out problems with the rivaling measure η_r: when a causally disconnected proposition is added to the effect, causal strength remains unchanged. Theorem 5 shows a analogous result for the case of causal prevention and the Good-Cheng measure η_{cg}, but in that case, it is less obvious that this property is harmful.

	Property										
Measure	FORM	EP	DM	CPS	WCPS	CE	SE	MUL	NDIE	NDIEP	CC
Pearl (η_p)	yes	yes	no	no	yes	yes	no	no	no	no	no
Probability Raise (η_{pr})	yes	yes	no	yes	yes	no	yes	no	no	no	no
Difference (η_d)	yes	no	yes	yes	yes	no	yes	yes	no	no	no
Probability Ratio (η_r, $\eta_{r'}$)	yes	no	yes	no	no	no	no	no	yes	yes	no
Cheng/Good (η_{cg})	yes	no	yes	yes	yes	no	no	no	no	yes	no
Conjunctive Closure (η_{cc})	yes	no	yes	no	no	no	no	no	no	no	yes

Table 4: A classification of different measures of causal strength according to the adequacy conditions they satisfy. FORM = Formality, EP = Effect Production, DM = Difference-Making, CPS = Causation-Prevention Symmetry, WCPS = Weak Causation-Prevention Symmetry, CE = Conditional Equivalence, SE = Separability of Effects, MUL = Multiplicativity on Single Paths, NDIE = No Dilution for Irrelevant Effects, NDIEP = No Dilution for Irrelevant Effects (Prevention), CC = Conjunctive Closure.

All in all, the above analysis provides good grounds for using η_d as a default measure of causal strength for purposes of prediction, intervention and control. Indeed, also Pearl (2001) bases his quantification of path-specific effects on η_d as underlying the baseline measure of causal strength. This paper provides an argument why Pearl's choice is actually sound, and why η_d is an adequate measure of average or total causal effect (Holland, 1986). The formal analysis also mirrors and supports practice- and decision-oriented arguments for η_d, as pointed out in Section 4.

What remains to do? First of all, I aim at generalizing the framework from binary 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. In many cases the analysis can even be applied without further modifications. For instance, imagine we are interested in a metric variable θ, and our observations E depend on the value of θ. The two hypotheses of interest are $H_0 : \theta \leq \theta_0$ and $H_1 : \theta > \theta_0$, a common scenario in statistics. Then, we can apply the above measures using $p_{\mathrm{H}_0}(\mathrm{E})$ and $p_{\mathrm{H}_1}(\mathrm{E})$ (and a Bayesian prior over H_0 and H_1) in order to determine the causal effect of H_0 vs. H_1 on E. The same can be said for the case that E is a real-valued variable: we can identify a subset I of the range of E that we are interested in, and quantify the causal effect of C on the proposition that $E \in I$. Thus, it is not always necessary that C and E be propositional variables. Eventually, our approach may go a longer way toward modeling causal inference in science than the binary framework suggests.

Second, the properties of the above measures 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)?

Third, 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 is likely to produce, the more specific the cause is to the effect. How does this concept relate to causal strength and

to what extent can both research programs learn from each other?

Fourth, one can apply this theory to canonical examples in the causation literature and explore whether these models 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 it is not difficult to come up with others. I hope that the results presented herein are promising enough to motivate a further pursuit of an axiomatic theory of causal strength.

A Proofs of the Theorems

Proof of Theorem 1: 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 $p(\mathrm{H}|\mathrm{E})$ as an adequate measure of degree of confirmation $c(\mathrm{E},\mathrm{H})$, 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 $p(\cdot)$, $c(\mathrm{E},\mathrm{H}) = f'(p(\mathrm{E}), p(\mathrm{H}), p(\mathrm{H} \wedge \mathrm{E}))$.

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

$$c(\mathrm{E}_1,\mathrm{H}) > c(\mathrm{E}_2,\mathrm{H}) \qquad \text{if and only if} \qquad p(\mathrm{H}|\mathrm{E}_1) \; > \; p(\mathrm{H}|\mathrm{E}_2).$$

Local Equivalence If H_1 and H_2 are logically equivalent given E, then $c(\mathrm{E},\mathrm{H}_1) = c(\mathrm{E},\mathrm{H}_2)$.

It is easy to see that Final Probability Incrementality translates into Effect Production when the pair $(\mathrm{H},\mathrm{E}_{1,2})$ is mapped to $(\mathrm{E},\mathrm{C}_{1,2})$:

$$\eta(\mathrm{C}_1,\mathrm{E}) > \eta(\mathrm{C}_2,\mathrm{E}) \qquad \text{if and only if} \qquad p(\mathrm{E}|\mathrm{C}_1) \; > \; p(\mathrm{E}|\mathrm{C}_2)$$

The same is true for Local Equivalence: with $(\mathrm{H}_{1,2},\mathrm{E})$ mapped to $(\mathrm{E}_{1,2},\mathrm{C})$. The condition postulates that if E_1 and E_2 are logically equivalent given C, then $\eta(\mathrm{C},\mathrm{E}_1) = \eta(\mathrm{C},\mathrm{E}_2)$. This is just the same as Conditional Equivalence, using different variable names.

Thus it remains to show that Formality (Causal Effect) 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 $\eta(\mathrm{C},\mathrm{E}) = f(p(\mathrm{C}), p(\mathrm{E}|do(\mathrm{C})), \mathrm{p}(\mathrm{E}|do(\neg\mathrm{C}))$.

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 assume that $p(\mathrm{C}) \in (0,1)$. This allows us to write

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

which we can transform into the equations

$$p(\mathrm{E}|do(\mathrm{C})) = \frac{p(\mathrm{E} \wedge \mathrm{C})}{p(\mathrm{C})} \qquad p(\mathrm{E}|do(\neg \mathrm{C})) = \frac{p(\mathrm{E}) - p(\mathrm{C})p(\mathrm{E}|do(\mathrm{C}))}{1 - p(\mathrm{C})} \tag{6}$$

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

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

f' is continuous because f and the functions in Equation (6) are. Thus we can extend f' canonically to the set $\{p(\mathrm{C}) \in \{0,1\}\}$. Hence we can invoke Schippers' theorem, and for any measure of causal strength η that satisfies Formality, Effect Production and Local Equivalence, $\eta(\mathrm{C}, \mathrm{E}) = p(\mathrm{E}|\mathrm{C})$ up to ordinal equivalence. q.e.d.

Proof of Theorem 2: Like in the previous theorem, we make use of a result for confirmation measures to prove the theorem (Crupi et al., 2013). We map Formality (Causal Strength) onto Formality (Confirmation) and Effect Production onto Final Probability Incrementality, like in the proof of the previous theorem. Finally, using a suitable change of variables ($\mathrm{C} \to \mathrm{E}$, $\mathrm{E}_1 \to \mathrm{H}_1$, $\mathrm{E}_2 \to \mathrm{H}_2$), it is easy to see that Separability of Effects is a stronger version of the following principle:

Disjunction of Alternative Hypotheses Assume H_1 and H_2 are inconsistent with

each other. Then,

$$
\begin{aligned}
c(\mathrm{H}_1 \vee \mathrm{H}_2, \mathrm{E}) > c(\mathrm{H}_1, \mathrm{E}) \qquad & \text{if and only if} \qquad p(\mathrm{H}_2|\mathrm{E}) > p(\mathrm{H}_2) \\
c(\mathrm{H}_1 \vee \mathrm{H}_2, \mathrm{E}) = c(\mathrm{H}_1, \mathrm{E}) \qquad & \text{if and only if} \qquad p(\mathrm{H}_2|\mathrm{E}) = p(\mathrm{H}_2) \\
c(\mathrm{H}_1 \vee \mathrm{H}_2, \mathrm{E}) < c(\mathrm{H}_1, \mathrm{E}) \qquad & \text{if and only if} \qquad p(\mathrm{H}_2|\mathrm{E}) < p(\mathrm{H}_2)
\end{aligned}
$$

Crupi et al. (2013) show that the difference measure $c_d := p(\mathrm{H}|\mathrm{E}) - p(\mathrm{H})$ is, up to ordinal equivalence, the only confirmation measure that satisfies Formality (Confirmation), Final Probability Incrementality and Disjunction of Alternative Hypotheses. Because these conditions are isomorphic to Formality (Causal Strength), Effect Production and Separability of Effects, their results prove the parallel result for measures of causal strength: the probability-raise measure $\eta_{pr}(\mathrm{C}, \mathrm{E}) = p(\mathrm{E}|do(\mathrm{C})) - p(\mathrm{E})$ is, up to ordinal equivalence, the only measure that satisfies the three latter conditions. q.e.d.

Proof of Theorem 3: The proof of this representation theorem proceeds in several steps. First, we will show that $\eta(\mathrm{C}, \mathrm{E}) = f(p(\mathrm{C}), p(\mathrm{E}|do(\mathrm{C})), \mathrm{p}(\mathrm{E}|do(\neg \mathrm{C}))$ does not depend on $p(\mathrm{C})$.

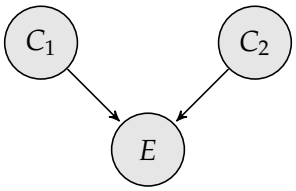

Figure 6: A classical collider/joint effect structure in a causal net.

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, C_1 and C_2 such that E is a joint effect of C_1 and C_2 with $x_1 = p(\mathrm{C}_1)$, $x_2 = p(\mathrm{C}_2)$, $y = p(\mathrm{E}|do(\mathrm{C}_1)) = p(\mathrm{E}|do(\mathrm{C}_2))$, $z = p(\mathrm{E}|do(\neg \mathrm{C}_1)) = p(\mathrm{E}|do(\neg \mathrm{C}_2))$. In this case, Difference-Making tells us that $\eta(\mathrm{C}_1, \mathrm{E}) = \eta(\mathrm{C}_2, \mathrm{E})$. However, on the other hand, we

also know

$$\eta(\mathrm{C}_1, \mathrm{E}) = f(x_1, y, z) \neq f(x_2, y, z) = \eta(\mathrm{C}_2, \mathrm{E})$$

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

The second step of the proof consists in deriving the equality

$$g(\alpha, \bar{\alpha}) \cdot g(\beta, \bar{\beta}) = g(\alpha\beta + (1-\alpha)\bar{\beta}, \bar{\alpha}\beta + (1-\bar{\alpha})\bar{\beta}) \tag{7}$$

To this end, recall the Bayesian network from the main paper. It is reproduced in Figure 7. Again, for the purpose of investigating the formal properties of g, we can focus on those cases where $p(\mathrm{E}|\pm\mathrm{C})$ and $p(\mathrm{E}|\pm\mathrm{C})$ agree.

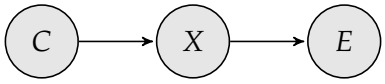

Figure 7: The Bayesian Network for causation along a single path.

We know by Multiciplity on Single Paths that

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

and at the same time,

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

Combining both equations yields

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

With the variable settings

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

equation (7) follows immediately.

Third, we are going to show that

$$g(x, y) = g(x - y, 0) \tag{8}$$

To this end, we first note a couple of facts about g:[8]

Fact 1 $g(\alpha, 0)g(\beta, 0) = g(\alpha\beta, 0)$. This follows immediately from equation (7) 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 $p(\mathrm{E}|do(\mathrm{C})) = 0$ and $p(\mathrm{E}|do(\neg\mathrm{C})) > 0$, in violation of Difference-Making. 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

[8] In the proof, negative arguments of g figure. This may look problematic, but it is not. We just show that any $g(\cdot, \cdot)$ that satisfies Equation (7) 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 $[0, 1]^2$.

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 (7) with $\alpha = 0$, $\bar{\alpha} = 1$.

These facts will allow us to derive Equation (8). Note that (8) is trivial if $y = 0$. So we can restrict ourselves to the case that $y > 0$. We choose the variable settings

$$\alpha = \frac{y - x}{y} \qquad \beta = 0$$
$$\bar{\alpha} = 0 \qquad \bar{\beta} = y$$

Then we obtain by means of Equation (7) 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

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

Hence, $\eta(\mathrm{C},\mathrm{E})$ is a function of $p(\mathrm{E}|do(\mathrm{C})) - \mathrm{p}(\mathrm{E}|do(\neg\mathrm{C}))$ only. From Difference-Making we can infer that g must be monotonically increasing in its first argument and monotonically decreasing in its second argument. This concludes the proof of Theorem 3. q.e.d.

Proof of Theorem 4: The proof relies on a move from the proof of Theorem 1 in Schupbach and Sprenger (2011). Consider three variables C, E_1 and E_2 which satisfy

the premises of the No Dilution for Irrelevant Effects Principle. This means that

$$\begin{aligned} p(\mathrm{E}_1 \wedge \mathrm{E}_2 | do(\mathrm{C})) &= p(\mathrm{E}_1 | do(\mathrm{C}))\, p(\mathrm{E}_2 | do(\mathrm{C})) \\ p(\mathrm{E}_1 \wedge \mathrm{E}_2 | do(\neg\mathrm{C})) &= p(\mathrm{E}_1 | do(\neg\mathrm{C}))\, p(\mathrm{E}_2 | do(\neg\mathrm{C})) \\ p(\mathrm{E}_2) &= p(\mathrm{E}_2 | do(\neg\mathrm{C})) \;=\; p(\mathrm{E}_2 | do(\mathrm{C})) \end{aligned}$$

In particular, it follows that

$$\begin{aligned} p(\mathrm{E}_1 \wedge \mathrm{E}_2 | do(\mathrm{C})) &= p(\mathrm{E}_2)\, p(\mathrm{E}_1 | do(\mathrm{C})) \\ p(\mathrm{E}_1 \wedge \mathrm{E}_2 | do(\neg\mathrm{C})) &= p(\mathrm{E}_2)\, p(\mathrm{E}_1 | do(\neg\mathrm{C})) \end{aligned}$$

According to Formality and Difference-Making, the causal strength measure η can be written as $\eta(\mathrm{C}, \mathrm{E}_1) = g(p(\mathrm{E}_1|do(\mathrm{C})), p(\mathrm{E}_1|do(\neg\mathrm{C})))$ for a continuous function g. From No Dilution for Irrelevant Effects and the above calculations we can infer that

$$\begin{aligned} g(p(\mathrm{E}_1|do(\mathrm{C})), p(\mathrm{E}_1|do(\neg\mathrm{C}))) &= \eta(\mathrm{C}, \mathrm{E}_1) \\ &= \eta(\mathrm{C}, \mathrm{E}_1 \wedge \mathrm{E}_2) \\ &= g(p(\mathrm{E}_1 \wedge \mathrm{E}_2|do(\mathrm{C})), p(\mathrm{E}_1 \wedge \mathrm{E}_2|do(\neg\mathrm{C}))) \\ &= g(p(\mathrm{E}_2)\, p(\mathrm{E}_1|do(\mathrm{C})), p(\mathrm{E}_2)\, p(\mathrm{E}_1|do(\neg\mathrm{C}))) \end{aligned}$$

Since we have made no assumptions on the values of these probabilities, we can infer the general relationship

$$g(x, y) = g(cx, cy). \tag{9}$$

for all $0 < c \leq \min(1/x, 1/y)$. Without loss of generality, let $x > y$. Then, choose c := 1/x. In this case, equation (9) becomes

$$g(x, y) = g(cx, cy) = g(1, y/x).$$

This implies that g must be a function of y/x only, that is, of the ratio

$p(\mathrm{E}|do(\mathrm{C}))/p(\mathrm{E}|do(\neg\mathrm{C}))$. Difference-Making then implies that all such functions must be monotonically increasing, concluding the proof of Theorem 4. q.e.d.

Proof of Theorem 5: We write the causal strength measure η as

$$\eta(\mathrm{C},\mathrm{E}) = \begin{cases} \eta^+(\mathrm{C},\mathrm{E}) & \text{for positive causation} \\ \eta^-(\mathrm{C},\mathrm{E}) & \text{for causal preemption} \end{cases}$$

We know from the previous theorem that $\eta^-(\mathrm{C},\mathrm{E})$ must be ordinally equivalent to $\eta_r(\mathrm{C},\mathrm{E})$. Now we show that all $\eta^+(\mathrm{C},\mathrm{E})$-measures are ordinally equivalent to $\eta_g(\mathrm{C},\mathrm{E}) = p(\neg\mathrm{E}|do(\neg\mathrm{C}))/p(\neg\mathrm{E}|do(\mathrm{C}))$. Since we have already shown that η_g and η_c are ordinally equivalent, this is sufficient for proving the theorem.

Because of Formality, we can represent η^+ by a function $g(x,y)$ with $x = p(\mathrm{E}|do(\mathrm{C}))$ and $y = p(\mathrm{E}|do(\neg\mathrm{C}))$. Suppose that there are $x > y$ and $x' > y' \in [0,1]$ such that $(1-y)/(1-x) = (1-y')/(1-x')$, but $g(x,y) \neq g(x',y')$. (Otherwise η^+ would just be a function of η_g, and we would be done.) In that case we can find a probability space such that $p(\mathrm{E}_1|do(\mathrm{C})) = x$, $p(\mathrm{E}_1|do(\neg\mathrm{C})) = y$, $p(\mathrm{E}_2|do(\mathrm{C})) = x'$, $p(\mathrm{E}_2|do(\neg\mathrm{C})) = y'$ and C screens off E_1 and E_2 (proof omitted, but straightforward). Hence $\eta^+(\mathrm{C},\mathrm{E}_1) \neq \eta^+(\mathrm{C},\mathrm{E}_2)$. By Weak Causation-Prevention Symmetry, we can then infer $\eta^-(\mathrm{C},\neg\mathrm{E}_1) \neq \eta^-(\mathrm{C},\neg\mathrm{E}_2)$.

However, since η^- is ordinally equivalent to η_r, there is a function f such that

$$\begin{aligned} \eta^-(\mathrm{C},\neg\mathrm{E}_1) &= f\left(\frac{p(\neg\mathrm{E}_1|do(\mathrm{C}))}{p(\neg\mathrm{E}_1|do(\neg\mathrm{C}))}\right) = f\left(\frac{1-x}{1-y}\right) \\ \eta^-(\mathrm{C},\neg\mathrm{E}_2) &= f\left(\frac{p(\neg\mathrm{E}_2|do(\mathrm{C}))}{p(\neg\mathrm{E}_2|do(\neg\mathrm{C}))}\right) = f\left(\frac{1-x'}{1-y'}\right) \end{aligned}$$

By assumption,

$$\frac{1-x}{1-y} = \left(\frac{1-y}{1-x}\right)^{-1} = \left(\frac{1-y'}{1-x'}\right)^{-1} = \frac{1-x'}{1-y'}$$

and so we can infer $\eta^-(\mathrm{C},\neg\mathrm{E}_1) = \eta^-(\mathrm{C},\neg\mathrm{E}_2)$, showing a contradiction.

Hence $\eta^+(\mathrm{C},\mathrm{E})$ can be represented by a monotonically increasing function of $p(\neg\mathrm{E}|do(\neg\mathrm{C}))/p(\neg\mathrm{E}|do(\mathrm{C}))$, completing the proof of Theorem 5. q.e.d.

Proof of Theorem 6: By Formality and Difference-Making, we have that $\eta(\mathrm{C},\mathrm{E}) = g(p(\mathrm{E}|do(\mathrm{C})), p(\mathrm{E}|do(\neg\mathrm{C})))$ for some continuous function $g : [0,1]^2 \to \mathbb{R}$. Assume that $\eta(\mathrm{C},\mathrm{E}_1) = \eta(\mathrm{C},\mathrm{E}_2) = t$, that C screens off E_1 and E_2 and that $p(\mathrm{E}_1|do(\mathrm{C})) = p(\mathrm{E}_2|do(\mathrm{C})) = x$, $p(\mathrm{E}_1|do(\neg\mathrm{C})) = p(\mathrm{E}_2|do(\neg\mathrm{C})) = y$, for some $x, y \in \mathbb{R}$. By the Conjunctive Closure Principle, we can infer

$$\eta(\mathrm{C},\mathrm{E}_1 \wedge \mathrm{E}_2) = \eta(\mathrm{C},\mathrm{E}_1) = g(x,y)$$

Moreover, we can infer

$$\begin{aligned}
\eta(\mathrm{C},\mathrm{E}_1 \wedge \mathrm{E}_2) &= g(p(\mathrm{E}_1 \wedge \mathrm{E}_2|do(\mathrm{C})), p(\mathrm{E}_1 \wedge \mathrm{E}_2|do(\neg\mathrm{C}))) \\
&= g(p(\mathrm{E}_1|do(\mathrm{C})) \cdot p(\mathrm{E}_2|do(\mathrm{C})), p(\mathrm{E}_1|do(\neg\mathrm{C})) \cdot p(\mathrm{E}_2|do(\neg\mathrm{C}))) \\
&= g(x^2,y^2)
\end{aligned}$$

Taking both calculations together, we obtain

$$g(x^2,y^2) = g(x,y) \tag{10}$$

as a structural requirement on the function g, since we have not made any assumptions on x and y.

Following Atkinson (2012) and his proof idea, we now define $u = \frac{\log x}{\log y}$ and define a function $f : \mathbb{R}^2 \to \mathbb{R}$ such that $f(x,u) := g(x,y)$. Equation (10) then implies the requirement

$$f(x^2,u) = g(x^2,y^2) = g(x,y) = f(x,u)$$

and by iterating the same procedure, we also obtain

$$f(x^{2n}, u) = f(x, u)$$

for some $n \in \mathbb{N}$. Due to the continuity of f, we can infer that f cannot depend on its first argument. Moreover, taking the limit $n \to \infty$ yields $f(x, u) = f(0, u)$. Hence, also

$$g(x, y) = f(0, u) = f(0, \log x / \log y)$$

and we see that

$$\eta(\mathrm{C}, \mathrm{E}) = h\left(\frac{\log p(\mathrm{E}|do(\mathrm{C}))}{\log p(\mathrm{E}|do(\neg\mathrm{C}))}\right)$$

for some function $h : \mathbb{R} \to \mathbb{R}$. It remains to show that h is monotonically increasing. Difference-Making implies that $\eta(\mathrm{C}, \mathrm{E})$ is an increasing function of $p(\mathrm{E}|do(\mathrm{C}))$ and a decreasing function of $p(\mathrm{E}|do(\neg\mathrm{C}))$. So it must be an increasing function of $\log p(\mathrm{E}|do(\mathrm{C})) / \log p(\mathrm{E}|do(\neg\mathrm{C}))$, too. This implies that h is a monotonically increasing function. Hence, all measures of causal strength that satisfy Formality, Difference-Making and the Conjunctive Closure Principle are ordinally equivalent to

$$\eta_{cc}(\mathrm{C}, \mathrm{E}) = \frac{\log p(\mathrm{E}|do(\mathrm{C}))}{\log p(\mathrm{E}|do(\neg\mathrm{C}))}. \qquad \text{q.e.d.}$$

References

Atkinson, D. (2012). Confirmation and Justification: A Commentary on Shogenji's Measure. *Synthese*, 184:49–61.

Beckers, S. (2016). *Actual Causation: Definitions and Principles*. PhD thesis, KU Leuven.

Bovens, L. and Hartmann, S. (2003). *Bayesian Epistemology*. Oxford University Press, New York.

Braham, M. and van Hees, M. (2009). Degrees of causation. *Erkenntnis*, 71:323–344.

Brössel, P. (2013). The problem of measure sensitivity redux. *Philosophy of Science*, 80(3):378–397.

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.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Lawrence & Erlbaum, Newark/NJ.

Cohen, M. P. (2016). Explanatory Justice: The Case of Disjunctive Explanations.

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.

Crupi, V. and Tentori, K. (2012). A Second Look at the Logic of Explanatory Power (with Two Novel Representation Theorems). *Philosophy of Science*, 79(3):365–385.

Crupi, V. and Tentori, K. (2013). Confirmation as partial entailment: A representation theorem in inductive logic. *Journal of Applied Logic*, 11:364–372.

Davies, H. T., Crombie, I. K., and Tavakoli, M. (1998). When can odds ratios mislead? *British Medical Journal*, 316(7136):989–991.

Deeks, J. (1998). When can odds ratios mislead? *British Medical Journal*, 317:1155.

Dowe, P. (2000). *Physical Causation*. Cambridge University Press, Cambridge.

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

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

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.

Hall, N. (2004). Two concepts of causation. In Collins, J., Hall, N., and Paul, L., editors, *Causation and Counterfactuals*, pages 255–276. MIT Press, Cambridge/MA.

Halpern, J. (2016). Sufficient Conditions for Causality to Be Transitive. *Philosophy of Science*.

Halpern, J. Y. and Hitchcock, C. (2015). Graded causation and defaults. *The British Journal for the Philosophy of Science*, 66:413–457.

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.

Hart, H. and Honoré, T. (1985). *Causation in the Law*. Oxford University Press, Oxford.

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.

Holland, P. W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*, 81:945–960.

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

Jensma, F. (2014). Marokkaanse afkomst heeft met criminaliteit niets van doen.

Kaiserman, A. (2016a). Causal Contribution. *Proceedings of the Aristotelian Society*.

Kaiserman, A. (2016b). Partial Liability.

King, N. B., Harper, S., and Young, M. E. (2012). Use of relative and absolute effect measures in reporting health inequalities: structured review. *British Medical Journal*, 345:5774.

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.

Meek, C. and Glymour, C. (1994). Conditioning and intervening. *The British Journal for the Philosophy of Science*, 45:1001–1021.

Milne, P. (1996). log[P(h/eb)/P(h/b)] is the One True Measure of Confirmation. *Philosophy of Science*, 63:21–26.

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.

Pearl, J. (2011). The structural theory of causation. In McKay Illari, P., Russo, F., and Williamson, J., editors, *Causality in the Sciences*. Oxford University Press, Oxford.

Reichenbach, H. (1956). *The Direction Of Time*. University, Berkeley and Los Angeles.

Rizzo, M. J. and Arnold, F. S. (1980). Causal Apportionment in Tort Law: An Economic Theory. *Columbia*, 85:1399–1429.

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.

Schupbach, J. N. and Sprenger, J. (2011). The Logic of Explanatory Power. *Philosophy of Science*, 78(1):105–127.

Shogenji, T. (2012). The Degree of Epistemic Justification and the Conjunction Fallacy. *Synthese*, 184:29–48.

Sistrom, C. L. and Garvan, C. W. (2004). Proportions, Odds, and Risk. *Radiology*, 230:12–19.

Skyrms, B. (1980). *Causal Necessity*. Yale University Press, New Haven and London.

Sloman, S. A. and Lagnado, D. (2015). Causality in Thought. *Annual Review of Psychology*, 66(1):223–247.

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

Sprenger, J. and Stegenga, J. (2016). Three Arguments for Absolute Outcome Measures. *Philosophy of Science*.

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.

Waldmann, M. R. and Hagmayer, Y. (2001). Estimating causal strength: The role of structural knowledge and processing effort. *Cognition*, 82:27–58.

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*.