**Foundations of a probabilistic theory of causal strength**

# Foundations of a Probabilistic Theory of Causal Strength*

Jan Sprenger†

November 10, 2017

**Abstract**

This paper develops axiomatic foundations for a probabilistic theory of causal strength as difference-making. I proceed in three steps: First, I motivate the choice of causal Bayes nets as an adequate 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 use these results to argue for a specific measure of causal strength: the difference that interventions on the cause make for the probability of the effect. I conclude by discussing my results and outlining future research avenues.

*Forthcoming in *The Philosophical Review*. The proofs of the theorems are contained in the supplemental material.

†Contact information: Department of Philosophy and Educational Sciences, and Center for Logic, Language and Cognition (LLC), Università degli Studi di Torino, Via Sant'Ottavio 20, 10124 Torino, Italy. Email: jan.sprenger@unito.it. Webpage: www.laeuferpaar.de

# Contents

# 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 their causal contributions to 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 since 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).

Principled proposals for explicating causal strength are rare and spread over different disciplines, each with their own motivation and intended context of application. This includes cognitive psychology (Cheng, 1997; Icard et al., 2017), computer science and machine learning (Pearl, 2000; Korb et al., 2009, 2011), statistics (Good, 1961a,b; Holland, 1986; Cohen, 1988), epidemiology and clinical medicine (Poole, 2010; Broadbent, 2013), 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; Kaiserman, 2017). Although these approaches use a common formalism—probability theory—, the proposed explications differ substantially (see the survey of Fitelson and Hitchcock, 2011). This may be due to the different purposes to which the measures are put: measuring predictive power, expressing counterfactual dependence, apportioning liability, and so on. The challenge for a philosophical theory of causal strength is to characterize the various measures and to evaluate whether we should prefer one

of them over its competitors, or whether we should use different measures in different contexts.

The paper proceeds as follows. Section 2 specifies the sense of causal strength explicated in this paper: the difference that causes make to their effects. I also motivate causal Bayes nets as an appropriate formal framework for this project. Section 3 derives representation theorems that characterize causal strength measures in terms of the adequacy conditions that they satisfy. These theorems lend support to preferring the difference measure $\eta_d(\mathrm{C},\mathrm{E}) = p(\mathrm{E}|\mathrm{C}) - p(\mathrm{E}|\neg\mathrm{C})$ over its competitors. Section 4 discusses possible objections while Section 5 sketches future research questions and concludes. All proofs are contained in the online appendix.

## 2 Interventions and Causal Bayes Nets

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 choose an economic policy: interventions increase the frequency of a particular response, but they do not guarantee it. Therefore, causal relevance is often explicated as **statistical relevance** or **probability-raising**: C is a cause of E if and only if C raises the probability of E (e.g., Reichenbach, 1956; Suppes, 1970; Cartwright, 1979; Eells, 1991). A cause is the more effective the more it raises the probability of an effect. Probability-raising captures the intuition that many causes make a difference to their effects without necessitating them.

It is well-known that purely probabilistic accounts of causality struggle to account for the asymmetry of causal relations. They dissolve the crucial difference between a causal inference (does bringing about C increase the probability of E?) and an observational inference (does learning C increase my confidence that E?). This is not the same: statistically associated variables, such as the number of ice cream sales and swimming pool visits on a particular day, need not be connected causally. More likely, they have a common cause, such as

temperatures and sunshine hours (cf. Reichenbach, 1956). Furthermore, unlike the cause-effect relation, statistical relevance is symmetric: if C raises the probability of E, then E also raises the probability of C.

Pearl (2000, 2011) notes that the problem is principled: 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 **interventionist account of causation** offers a principled solution to this problem. The idea behind probability-raising is modified to the effect that a variable $C$ is a cause of another variable $E$ if and only if an intervention on $C$ changes the probability that $E$ takes a particular value.[1] Because the intervention breaks the influence of the other causes of $C$, it removes spurious correlations. After intervening on the number of swimming pool visitors (e.g., by closing the pool for renovation works), learning the number of visitors (zero) does not tell us anything about temperatures or ice cream sales. By now, the interventionist account of causation is prevalent in philosophical discussions of causality (Meek and Glymour, 1994; Woodward, 2003, 2012) as well as in scientific applications such as causal search and discovery algorithms (Pearl, 2000; Spirtes et al., 2000).

For interventionists, causal reasoning is relative to the choice of a causal model $M$: a set of variables with a joint probability distribution and specific causal dependencies. The latter are represented by a directed acyclical graphs (DAG) $G$, consisting of a set of vertices (=variables) and directed edges (e.g., Figure 1). DAGs codify Pearl's "causal assumptions" underneath our causal reasoning. 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 endowed with a joint probability distribution over the variables are called **causal Bayes nets**. In a DAG, an intervention on $C$

[1] The probability function $p(\cdot)$ can be interpreted objectively (frequencies, propensities, best-system chances) or as subjective degrees of belief, dependent on the context.

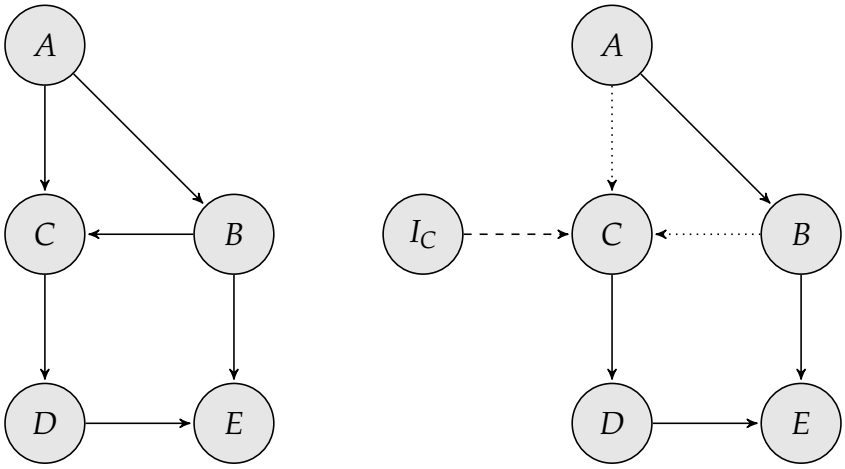


Figure 1: Two DAGs with and without an intervention on the cause $C$. The arrows leading into $C$ (dotted) are disrupted by the intervention, the intervention itself is represented by a dashed arrow.

amounts to breaking all arrows that lead into $C$, and to control the value that $C$ takes. This can consist in setting $C$ to a particular value, but also in imposing a specific probability distribution on $C$ (e.g., in medical trials, patients are randomly assigned to treatment and control groups).

The two graphs in Figure 1 show a causal graph before and after a (hypothetical) intervention on $C$. I denote variables by italic letters (e.g., $C$) and use regular roman letters for particular values they take (e.g., $\mathrm{C}, \neg\mathrm{C}, \mathrm{C}'$)—see Bovens and Hartmann (2003). The intervention on $C$ is represented by the node $I_C$. Activating $I_C$ controls the value of $C$ and removes the influence of the parent nodes $A$ and $B$. Intervening on $C$ leads to an augmented causal model $M^*$ with modified DAG $G^*$ and probability distribution $p^*$, represented by the right graph of Figure 1. Since intervening on $C$ does not alter the way $C$ acts on causally downstream variables, we set $p^*(\cdot|\mathrm{C}) = p(\cdot|\mathrm{C})$ for all values of $C$.

This paper combines the probabilistic and the interventionist perspective. I measure the causal strength of C for a target effect E as the degree to which C makes a difference to E in the post-intervention distribution $p^*$. This approach preserves the asymmetry of causal relations: if we had intervened on $E$ instead of $C$, all arrows into $E$ in Figure 1 would have been cut, making $C$ and $E$ statistically independent. Causal strength between C and E would have been

nil. Moreover, by cutting the arrows that lead into $C$, the intervention removes possible spurious correlations between $C$ and $E$ due to their common causes $A$ and $B$.[2]

I now add some precision to these ideas. A variable $C = (\Omega_C, \mathcal{C})$ in a causal Bayes net corresponds to a set of possible values $\Omega_C$ and a $\sigma$-algebra $\mathcal{C}$: a set of subsets of $\Omega_C$ which contains $\Omega_C$ itself and is closed under natural set-theoretic operations such as union and complement. For instance, if $C$ is a real-valued variable ($\Omega_C = \mathbb{R}$), then a particularly "natural" $\sigma$-algebra for $\Omega_C$ is the Borel $\sigma$-algebra: it contains single numbers such as $\{2\}$ or $\{3\}$, but also intervals such as $[2, 3)$, $(-1, 1]$ or $[10, \infty)$. In the rest of the paper, we often replace the set-theoretic formalism by propositions about the values of $C$. It is more intuitive and the mathematical structures (e.g., the associated $\sigma$-algebras) are fully isomorphic.

A **causal model** $M = (G, p(\cdot))$ of the interaction between $C$ and $E$ is a directed acyclical graph $G$ which includes, among others, the variables $C = (\Omega_C, \mathcal{C})$ and $E = (\Omega_E, \mathcal{E})$. It also contains a probability distribution $p(\cdot)$ over the variables in $G$. We can also define the post-intervention model $M^* = (G^*, p^*(\cdot))$ that emerges from $M$ by intervening on $C$, and cutting all edges that run into $C$. Causal strength depends on the features of the model $M^* = (G^*, p^*(\cdot))$ that emerges from $M$ by intervening on $C$, and in particular on the joint probability distribution over $C$ and $E$. A causal strength measure maps, in essence, elements of $\mathcal{C} \times \mathcal{E}$ (e.g, the pair (C, E)) to a real number, denoted by $\eta(C, E)$.[3] More precisely, we explicate causal strength contrastively:

[2]Technical details of this approach are discussed in Korb et al. (2009) and Korb et al. (2011). They also discuss the problem of non-causal paths between $C$ and $E$ via known common effects. Such paths introduce a non-causal probabilistic dependency between $C$ and $E$. Following their suggestion, I deactivate all non-causal paths between $C$ and $E$: causal strength between C and E should not be affected by whatever values their joint effects take. For evaluating the effects of university education on salary, we should not condition on joint effects such as driving a Porsche or becoming a U.S. senator.

[3]This definition shows why it was necessary to introduce the technical concept of a $\sigma$-algebra: we do not only want to talk about causal strength between variables that take particular values, but also about more general features such as a variable exceeding a particular threshold. For instance, we can now determine the causal effect that temperatures greater than 30 °C have on the occurrence of thunderstorms in the afternoon.

as a function of how C, as opposed to another value of $C$, affects the post-intervention probability of E. This brings us to the following very general adequacy constraint (reference to $M^*$ suppressed for the sake of convenience):

**Generalized Difference-Making** For a putative cause $\mathrm{C} \in \mathcal{C}$ and a putative effect $\mathrm{E} \in \mathcal{E}$, there exists a $\mathrm{C}' \in \mathcal{C}$ and a real-valued, continuous function $f : [0,1]^2 \to \mathbb{R}$ such that for the causal strength of C on E, $\eta(C, E)$:

$$\eta(\mathrm{C}, \mathrm{E}) = f(p^*(\mathrm{E}|\mathrm{C}), p^*(\mathrm{E}|\mathrm{C}'))$$

where $f$ is non-decreasing in the first argument and non-increasing in the second argument.

The idea of causal strength as difference-making is an intuition shared by counterfactual, probabilistic and interventionist accounts of causation alike. Causal strength is the higher, the more probable E is given C, and the less probable E is given the contrastive value C′. How C′ should be chosen is a matter to which I will get back in the next section.

Generalized Difference-Making quantifies causal strength with respect to a particular causal model. Similarly, it focuses on a **single background context**, sidestepping a substantial discussion in the field of probabilistic causation (e.g., Cartwright, 1979; Dupré, 1984; Eells, 1991). This makes intuitive sense: when we investigate the relationship between beer consumption and obesity, causal strength depends on characteristics of the population such as age, dietary habits and general lifestyle. Compared to couch potatoes, active athletes are less likely to gain weight when they drink two or three pints a day. Causal strength claims are always relative to such a choice of context, symbolically represented by other variables that have an effect on the probability of E (e.g., $A$ and $B$ in the DAG of Figure 1). Similarly, causal strength depends on the levels of the cause variable that we compare (i.e., three vs. two, or five vs. zero pints per day). This choice is codified in the post-intervention distribution $p^*(\cdot)$. By contrast, I do not include external factors such as typicality, defaults and normative expectations, which have been argued to affect causal

judgments (Knobe and Fraser, 2008; Hitchcock and Knobe, 2009; Halpern and Hitchcock, 2015). I discuss objections to my approach in Section 4.

Notably, causal strength is blind to the presence of multiple paths leading from $C$ to $E$, or the number of mediators between $C$ and $E$ (see Figure 2). This choice is conscious. Mediating variables 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. $\eta(\mathrm{C}, \mathrm{E})$ amalgamates the effects of C on E via different paths into one number—the **total effect of C on E** (e.g., Dupré, 1984; Eells, 1991). This does not rule out a path-specific perspective, quite to the contrary. Measures of path-specific effect supervene on elementary measures of causal strength that quantify causal strength between adjacent variables (e.g., Pearl, 2001). In this sense, this paper lays the foundations for path-specific analyses of causal strength.

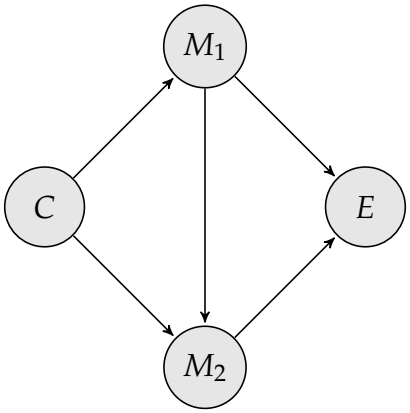


Figure 2: Mediators on the paths linking cause C and effect E.

# 3 Probabilistic Measures of Causal Strength

Generalized Difference-Making does not specify how $C'$ should be chosen and how $p^*(\mathrm{E}|\mathrm{C})$ and $p^*(\mathrm{E}|\mathrm{C}')$ should be combined. This leaves open a number of ways to explicate causal strength. Some candidate measures of causal strength that align with Generalized Difference-Making are surveyed in Fitelson and Hitchcock (2011) and recapitulated in Table 1.

How should we deal with this plurality of measures of causal strength? Two major attitudes are possible. First, there is **monism**: there is only one adequate measure (or equivalence class of measures) of causal strength. Second, there is **pluralism**: no single measure satisfies all the conditions that an adequate measure of causal strength should possess. This is perhaps the default view. After all, intuitions about complex concepts such as causal strength may pull into different directions and lead to a set of adequacy conditions which a single measure cannot possibly satisfy. This is at least the lesson one might draw from the analogous projects of finding a probabilistic explication of evidential support, or the coherence of a set of propositions (e.g., Fitelson, 1999; Meijs, 2005; Brössel, 2013; Crupi, 2013).

I contend that the prospects for causal strength monism are more promising. This monism is *qualified*: it is based on understanding causal strength as counterfactual difference-making, as informing our expectations on the efficacy of interventions on $C$.[4] Within the explicative framework outlined by Generalized Difference-Making, there is a single adequate measure of causal strength, namely $\eta_d(\mathrm{C},\mathrm{E}) = p^*(\mathrm{E}|\mathrm{C}) - p^*(\mathrm{E}|\neg\mathrm{C})$ (Eells, 1991; Pearl, 2001).

In what follows, I present two different constructive arguments in favor of $\eta_d$ and a negative argument against probability ratio measures. The arguments establish that the prospects for monism in measuring causal strength are brighter than in other debates, such as measuring coherence, confirmation and explanatory power. **Ordinally equivalent** measures—that is, measures that im-

[4]Whether this monism transfers to measuring causal strength as degrees of actual causation or "cause in fact" (Halpern and Pearl, 2005a,b)—that is, as a basis for attribution and responsibility—is outside the scope of this paper. Icard et al. (2017) investigate this question by a combination of theoretical and empirical methods.

| | |
|---|---|
| Pearl (2000) | $\eta_{ph}(\mathrm{C},\mathrm{E}) = p(\mathrm{E}\vert\mathrm{C})$ |
| Suppes (1970) | $\eta_{pr}(\mathrm{C},\mathrm{E}) = p(\mathrm{E}\vert\mathrm{C}) - p(\mathrm{E})$ |
| Eells (1991) | $\eta_{d}(\mathrm{C},\mathrm{E}) = p(\mathrm{E}\vert\mathrm{C}) - p(\mathrm{E}\vert\neg\mathrm{C})$ |
| "Galton" (covariation) | $\eta_{ga}(\mathrm{C},\mathrm{E}) = 4p(\mathrm{C})\,p(\neg\mathrm{C})[p(\mathrm{E}\vert\mathrm{C}) - p(\mathrm{E}\vert\neg\mathrm{C})]$ |
| Lewis (1986) | $\eta_{r}(\mathrm{C},\mathrm{E}) = \dfrac{p^*(\mathrm{E}\vert\mathrm{C})}{p^*(\mathrm{E}\vert\neg\mathrm{C})}$ |
| Cheng (1997) | $\eta_{c}(\mathrm{C},\mathrm{E}) = \dfrac{p(\mathrm{E}\vert\mathrm{C}) - p(\mathrm{E}\vert\neg\mathrm{C})}{1 - p(\mathrm{E}\vert\neg\mathrm{C})}$ |
| Good (1961a,b) | $\eta_{g}(\mathrm{C},\mathrm{E}) = \log\dfrac{1 - p(\mathrm{E}\vert\neg\mathrm{C})}{1 - p(\mathrm{E}\vert\mathrm{C})}$ |

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

pose the same causal strength rankings on any set of cause-effect pairs—will be identified with each other. Formally, two measures $\eta$ and $\eta'$ are ordinally equivalent if and only if for all cause-effect pairs $(\mathrm{C}_1,\mathrm{E}_1)$ and $(\mathrm{C}_2,\mathrm{E}_2)$ in a causal model $M$,

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

Ordinally equivalent measures can be represented as monotonically increasing functions of each other. Typical cases are addition or multiplication of a constant, or rescalings of the type $\eta' = \log\eta$. In other words, ordinally equivalent measures may use different scales, but they agree in all comparative judgments and share most philosophically interesting properties.

The following subsections provide representation theorems for measures of causal strength and use these theorems to buttress normative arguments for a particular measure (up to ordinal equivalence). The representation theorems have independent value, too: they allow to translate any normative evaluation of the adequacy conditions—also an evaluation that is fundamentally differ-

ent from the one I propose—into a corresponding ranking of causal strength measures.

## 3.1 Argument 1: Separability and Difference Measures

Suppose we examine how an intervention on a class of students, such as increasing the assignment load (C), affects their exam results. The result variable $R$ can take three values: pass with honors ($R_1$), regular pass ($R_2$) and fail ($R_3$). Suppose we know the causal strength of more assignments for passing with honors (i.e., $\eta(C, R_1)$) and also the causal strength of more assignments for regular passes (i.e., $\eta(C, R_2)$). Since passing is just the disjunction of regular pass and pass with honors, the causal strength of C for $R_1 \vee R_2$ should exceed the causal strength for both $R_1$ and $R_2$ only if both are indeed caused rather than prevented by C. In other cases, causal strength for the aggregate effect should be down. This is equivalent to the following: causal strength increases under adding a disjunct to the effect when the cause is positively relevant to the disjunct, and decreases when it is negatively relevant. We obtain the following adequacy condition:

**Separability of Effects** For $C \in \mathcal{C}$ and mutually exclusive $E, E' \in \mathcal{E}$:

$$\begin{aligned}
\eta(C, E \vee E') > \eta(C, E) \quad & \text{if and only if} \quad p^*(E'|C) > p^*(E'|C') \\
\eta(C, E \vee E') = \eta(C, E) \quad & \text{if and only if} \quad p^*(E'|C) = p^*(E'|C') \\
\eta(C, E \vee E') < \eta(C, E) \quad & \text{if and only if} \quad p^*(E'|C) < p^*(E'|C')
\end{aligned}$$

From Separability of Effects and Generalized Difference-Making, it is possible to prove the following representation theorem:

**Theorem 1 (Representation Theorem for Difference Measures)** *All measures of causal strength that satisfy Separability of Effects and Generalized Difference-Making are of the form*

$$\eta(C, E) = p^*(E|C) - p^*(E|C').$$

This theorem implies that $\eta(\mathrm{C}, \mathrm{E})$ must be the difference of the rate of E under C and a relevant contrast class C′. All such measures satisfy the equality

$$\eta(\mathrm{C}, \mathrm{E} \vee \mathrm{E}') = \eta(\mathrm{C}, \mathrm{E}) + \eta(\mathrm{C}, \mathrm{E}')$$

for mutually exclusive E and E′, allowing for an easy computation of aggregate causal strength from causal strength of the disjuncts.

For choosing C′, there are two particularly natural candidates. First, the choice $\mathrm{C}' = \Omega_C$ (=no restrictions on the value of $C$) leads to a measure that quantifies how much C raises the "natural" occurrence rate of E (cf. Pearl, 2011, 717).

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

Second, we can choose $\mathrm{C}' = \neg\mathrm{C}$ and measure the difference between the presence and absence of C (Eells, 1991):

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

This measure captures the degree to which E depends on C. For instance, in a randomized controlled trial (RCT) where we compare two levels of a drug, $\eta_d$ quantifies the difference in incident rates between the treatment and the control group.

While both measures are natural and frequently cited candidates for measuring causal strength, there is a clear argument for preferring $\eta_d$. We apply causal strength in contexts where we intervene, or could hypothetically intervene on the cause. Both measures depend, to some extent, on the post-intervention probability distribution of $C$—and in particular, on the relative frequency of the alternative values to C—say, $\mathrm{C}_1, \mathrm{C}_2$ and $\mathrm{C}_3$. This dependency is not problematic because it expresses the relevant contrast class (e.g., do we compare the efficacy of a new drug to a placebo, to the previous standard treatment, or to a mixture of both?).

However, in addition to this, $\eta_{pr}$ introduces a strong dependence on the base rate of C because $p^*(\mathrm{E}) = p^*(\mathrm{C})p^*(\mathrm{E}|\mathrm{C}) + p^*(\neg\mathrm{C})p^*(\mathrm{E}|\neg\mathrm{C})$. This consequence is hardly acceptable. Causal strength should not depend on pragmat-

ically motivated decisions expressed by the post-intervention frequency of C and ¬C, such as the number of patients that we allocate to the treatment and control group. In particular, causal strength in a treatment-control experiment should not be highly sensitive to whether the treatment group consists of 100 participants and the (possibly heterogenous) control group of 50 participants, or the other way round. Both experiments should allow for the same kind of causal strength inferences, but for

$$\eta_{pr}(\mathrm{C},\mathrm{E}) = p^*(\mathrm{E}|\mathrm{C}) - p^*(\mathrm{E}) = (1 - p^*(\mathrm{C})) \cdot (p^*(\mathrm{E}|\mathrm{C}) - p^*(\mathrm{E}|\neg\mathrm{C}))$$

the range of possible causal strength values is $[-1/3, 1/3]$ in the first case and $[-2/3, 2/3]$ in the second case. This is clearly an undesirable consequence.[5] Since these arguments pertain to the choice of C′ and can be generalized beyond the particular function that combines $p^*(\mathrm{E}|\mathrm{C})$ and $p^*(\mathrm{E}|\mathrm{C}')$, $\mathrm{C}' = \neg\mathrm{C}$ will be a default assumption in the remainder of the paper.

Note that both measures satisfy two important causation-prevention symmetries which will be important later on. I follow Fitelson and Hitchcock (2011) in explicating the degree to which C prevents E as the degree to which C causes ¬E, that is, the absence of E. To be able to measure causation and prevention on the same scale, we demand that the (preventive) causal strength of C for ¬E is the negative of the causal strength of C for E.

**Causation-Prevention Symmetry (CPS)**

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

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, $\eta(\mathrm{C},\mathrm{E}) = 0$ denotes neutral causal strength.[6] Those who share the basic intuition behind the Causation-

[5]The proposed argument is typical of causation as difference-making. For questions of causal attribution that arise in the debate about actual causation, empirical evidence suggests that we judge statistically abnormal causes to be stronger than statistically normal causes (Kominsky et al., 2015; Icard et al., 2017).

[6]This should not be conflated with causal irrelevance. A cause can be relevant for an effect,

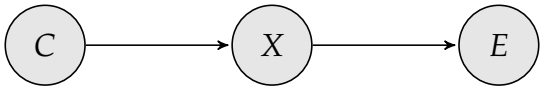


Figure 3: A DAG representing causation along a single path.

Prevention Symmetry, but do not want to subscribe to particular scaling properties, can still accept the following purely ordinal version of the symmetry: if C is a stronger cause of $E_1$ than of $E_2$, then C prevents $\neg E_1$ more than $\neg E_2$.

**Weak Causation-Prevention Symmetry (WCPS)**

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

## 3.2 Argument 2: The Multiplicativity Principle

How should causal strength combine on the single path of Figure 3? If causal strength is the ability of the cause to make a difference to the effect, then 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 for an intermediate cause X, $\eta(C, E) = g(\eta(C, X), \eta(X, E))$?

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 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: causal strength will still be weak. Similarly, if both links are weak, the overall link will be even weaker. On the other hand, if one link is maximally strong (e.g., $\eta(C, X) = 1$), then the strength of the entire chain will just be the strength of the rest of the chain. Perfect connections between two nodes neither raise nor attenuate overall causal strength (see also Good, 1961a, 311–312).

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.

A very simple operator that satisfies all these requirements is multiplication. This suggests the following principle:

**Multiplicativity** If the variables $C$ and $E$ are connected via a single path with intermediate node $X$, then for $\mathrm{C} \in \mathcal{C}$, $\mathrm{E} \in \mathcal{E}$, and $\mathrm{X} \in \mathcal{X}$:

$$\eta(\mathrm{C}, \mathrm{E}) = \eta(\mathrm{C}, \mathrm{X}) \cdot \eta(\mathrm{X}, \mathrm{E})$$

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

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

Multiplication may not be the only operator that fits the bill. However, it is clearly the simplest one and ceteris paribus, simplicity is an added benefit for an explicatum (Carnap, 1950, 5). The simple mathematical form contributes to theoretical fruitfulness, as we see in the above equation for longer causal chains.

Second, multiplicativity agrees with a lot of scientific practice. Suppose there is a linear dependency between variables $E$ and $X$, modeled by the equation $E = \alpha X + i$. In those cases, the regression coefficient $\alpha$ is commonly interpreted as indicating the size of the causal effect that $X$ has on $E$. When $X$ depends linearly on $C$, too (e.g., $X = \beta C + i'$), the relation between $C$ and $E$ reads $E = \alpha\beta C + \alpha i' + i$, and the regression coefficient between both variables is equal to $\alpha\beta$—in agreement with Multiplicativity.[7]

Third, suppose that in the absence of C, it is very unlikely that E: $p^*(\mathrm{E}|\neg\mathrm{C}) \approx 0$. In such circumstances, causal strength is the higher the more likely C is to bring about E. Modeling causal strength as a linear function of $p^*(\mathrm{E}|\mathrm{C})$, up to a given degree of precision, is particularly intuitive. I call this the Proportionality Principle.[8] It is not difficult to prove that Multiplicativity

[7] Incidentally, this condition assumes transitivity of causation, which is warranted for $\mathrm{C}' = \neg\mathrm{C}$ (Korb et al., 2011; Halpern, 2016).

[8] Formally, Proportionality can be defined as follows. Suppose $f : [0,1]^2 \to \mathbb{R}$ is an analytic function that represents $\eta$ mathematically, in agreement with Generalized Difference Making.

holds up to arbitrary degree of precision when Proportionality holds. Hence, Multiplicativity also flows from a natural way to think about the strength of necessary causes. Indeed, if C is necessary for X, and X is necessary for E, then C is also necessary for E and the equation $p^*(\mathrm{E}|\mathrm{C}) = p^*(\mathrm{E}|\mathrm{X}) \cdot p^*(\mathrm{X}|\mathrm{C})$ holds, vindicating Proportionality (and Multiplicativity). These three independent arguments validate Multiplicativity as a simple, attractive and conceptually sound principle.

We can now characterize all measures that satisfy Multiplicativity alongside Generalized Difference-Making:

**Theorem 2 (Representation Theorem for $\eta_d$)** *All measures of causal strength that satisfy Generalized Difference-Making with the contrast class $\mathrm{C}' = \neg\mathrm{C}$ and Multiplicativity are ordinally equivalent to*

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

The probability difference is a simple and intuitive quantity that measures causal strength by comparing the probability that different interventions on C impose on E. Indeed, $\eta_d$ 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 (see Section 3.4). Holland (1986, 947) calls $\eta_d$ the "average causal effect" of C on E—a label that is motivated by the fact that $\eta_d$ aggregates the strength of different causal links. Pearl (2001) uses $\eta_d$ as the basis for developing a path-sensitive theory of causal strength.

Finally, $\eta_d$ can be written as $\eta_d(\mathrm{C}, \mathrm{E}) = p^*(\mathrm{E}|\mathrm{C}) + p^*(\neg\mathrm{E}|\neg\mathrm{C}) - 1$. In this representation, causal strength depends linearly on two salient quantities: $p^*(\mathrm{E}|\mathrm{C})$ and $p^*(\neg\mathrm{E}|\neg\mathrm{C})$. They express the probability that C is *sufficient* for E and the probability that C is *necessary* for E (see also Pearl, 2000). $\eta_d$ shares this property with Icard et al.'s (2017) measure of actual causation.

---

Then,

$$\forall\varepsilon > 0\, \exists\delta > 0 :\ \forall\alpha > 0,\, 0 < x < 1,\, y < \delta : \qquad |f(\alpha x, y) - \alpha f(x, y)| < \varepsilon$$

Proving Multiplicativity from Proportionality is a matter of straightforward calculus.

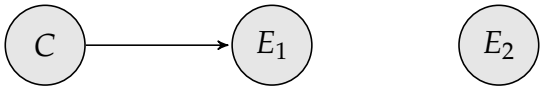


Figure 4: An effect $E_2$ which is irrelevant regarding the causal relation between $C$ and $E_1$.

## 3.3 Argument 3: Dilution and the 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, $\mathrm{E}_2$) is independent of C and of $\mathrm{E}_1$? In such circumstances, we may call $\mathrm{E}_2$ 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 $\mathrm{E}_1$ to $\mathrm{E}_1 \wedge \mathrm{E}_2$, or it is not. **Dilution** means that adding $\mathrm{E}_2$ to $\mathrm{E}_1$ 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 $\mathrm{C} \in \mathcal{C}, \mathrm{E}_1 \in \mathcal{E}_1, \mathrm{E}_2 \in \mathcal{E}_2$, let $E_2 \perp\!\!\!\perp C$, $E_2 \perp\!\!\!\perp E_1$ 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.[9] In fact, they are all ordinally equivalent to the probability ratio measure (Lewis, 1986), as the following theorem demonstrates.

**Theorem 3 (Representation Theorem for $\eta_r$ and $\eta_{r'}$)** *All measures of causal strength that satisfy Generalized Difference-Making with the contrast class $\mathrm{C}' = \neg\mathrm{C}$ and No Dilution for Irrelevant Effects are ordinally equivalent to*

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

[9] 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 $C$.

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

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

This result can be interpreted as a *reductio ad absurdum* of probability ratio measures. 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 $\mathrm{E}_1$. Rain in New York on November 26, 2016 (C) affects umbrellas sales in that city ($\mathrm{E}_1$), but it does not affect whether FC Barcelona will win their next Champions League match ($\mathrm{E}_2$). 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 $\eta_r$ and $\eta_{r'}$.

The problems extend beyond the class of probability ratio measures. Consider the following restriction of No Dilution to the class of causal prevention:

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

Together with the Weak Causation-Prevention Symmetry, this adequacy condition is sufficient to single out a particular class of measures:

**Theorem 4 (Representation Theorem for $\eta_{cg}$)** *All measures of causal strength that satisfy Generalized Difference-Making with the contrast class* $\mathrm{C}' = \neg\mathrm{C}$*, No Dilution for Irrelevant Effects (Prevention) and Weak Causation-Prevention Symmetry are ordinally equivalent to*

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

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

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

(that is, $\eta_{cg}$ without the above case distinction) from theoretical considerations about how agents perform causal induction and called it the "causal power" of C on E. Cheng's measure is in turn ordinally equivalent to the measure

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

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.

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

The two previous theorems elucidate that $\eta_r$ and $\eta_{cg}$ are based on the same principle: No Dilution for Irrelevant Effects. Since this property is highly suspicious, the representation results also provide evidence against $\eta_{cg}$ and its cognates $\eta_c$ and $\eta_g$, ruling out a prima facie attractive class of alternative measures.

## 3.4 Application: Quantifying Causal Effect in Medicine

A classical case of measuring causal strength concerns Randomized Controlled Trials (RCTs) in medicine. The various outcome measures can be translated into our framework by writing observed relative frequencies of certain events as conditional probabilities under the different levels of the cause (i.e., the treatment level). For example:

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

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

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

It is not difficult to relate these measures to measures of causal strength. For example, RR is just the familiar probability ratio measure $\eta_r$, whereas ARR turns out to be the difference measure $\eta_d$. RRR $=$ RR $-1$ is ordinally equivalent to $\eta_r$.

Normative arguments in favor or against causal strength measures carry over to these effect size measures. Since the probability ratio measure $\eta_r$ 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. The causal effect of a painstiller on relieving headache is, according to $\eta_r$, as big as the causal effect of that drug on relieving headache *and* a completely unrelated symptom, e.g., lowering cholesterol levels. $\eta_r$ grossly misrepresents causal relevance: it conceals that the high causal strength of the drug for both symptoms taken together is exclusively due to its effect on pain relief. Doctors may be misled into prescribing the drug for lowering cholesterol levels, even if it is ineffective for that purpose.

On the other hand, the defining properties of $\eta_d$, such as combining causal strength along a single path with the formula $\eta_d(\mathrm{C},\mathrm{E}) = \eta_d(\mathrm{C},\mathrm{X}) \cdot \eta_d(\mathrm{X},\mathrm{E})$, suit clinical practice very well. For example, doctors can see that overall causal strength must be weak if one of the links is tenuous. These theoretical features nicely square with decision-theoretic and epistemic arguments for preferring absolute over relative risk measures in medicine, such as the neglect of base rates in relative risk measures, and the sufficiency of $\eta_d$ for identifying the most promising treatment (Stegenga, 2015; Sprenger and Stegenga, 2017). Briefly, the scientific application confirms our theoretical diagnosis: $\eta_d$ is superior to $\eta_r$ and other probability ratio measures.

# 4 Discussion of the Results

Let us take stock. The previous section has provided three independent arguments for regarding $\eta_d$ as a default measure of causal strength. The first argument was based on the Separability of Effects property, the second on the

Multiplicativity Principle, and the third on the No Dilution Principle. The first two arguments showed that $\eta_d$ is the only measure that satisfies those desirable properties. The third argument, by contrast, points out problems with the probability ratio family and related measures ($\eta_r$, $\eta_{r'}$, $\eta_{cg}$, $\eta_c$, $\eta_g$) based on the No Dilution property.

Each individual argument makes a good case. Cumulatively, things look even better since the three arguments operate independently from each other. Still, one may have principal doubts about uniqueness claims for causal strength measures. I will now play *advocatus diaboli* and introduce two measures that neither have the attractive properties of $\eta_d$ nor the problematic properties of the No Dilution measures (e.g., $\eta_r$, $\eta_{cg}$).

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 $t$. One may want to say that the overall side effect of the medical drug is also equal to $t$ since there is no interaction between both effects.

**Conjunctive Closure** For $\mathrm{C} \in \mathcal{C}, \mathrm{E}_1 \in \mathcal{E}_1$ and $\mathrm{E}_2 \in \mathcal{E}_2$, with $E_1 \perp\!\!\!\perp E_2$ conditional on $C$, the following implication holds:

$$\eta(\mathrm{C}, \mathrm{E}_1) = \eta(\mathrm{C}, \mathrm{E}_2) = t \qquad \Rightarrow \quad \eta(\mathrm{C}, \mathrm{E}_1 \wedge \mathrm{E}_2) = t \tag{2}$$

This principle facilitates calculations because we can now infer the strength of a cause C for an aggregate effect from the strength of C for the individual effects. Measures that satisfy Conjunctive Closure can be characterized neatly (Atkinson, 2012):

**Theorem 5 (Representation Theorem for $\eta_{cc}$)** *All measures of causal strength that satisfy Generalized Difference-Making with the contrast class $C' = \neg\mathrm{C}$ and Conjunctive Closure are ordinally equivalent to*

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

Although this measure fails to satisfy Separability of Effects and Multi-

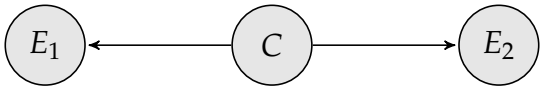


Figure 5: A DAG representing a common cause structure.

plicativity, it may be a reasonable measure in contexts where we would like to quantify the **average strength** of a cause for a variety of independent effects. See Figure 5 for an illustration. For that causal structure, it is always the case that

$$\min(\eta_{cc}(\mathrm{C}, \mathrm{E}_1), \eta_{cc}(\mathrm{C}, \mathrm{E}_2)) \leq \eta_{cc}(\mathrm{C}, \mathrm{E}_1 \wedge \mathrm{E}_2) \leq \max(\eta_{cc}(\mathrm{C}, \mathrm{E}_1), \eta_{cc}(\mathrm{C}, \mathrm{E}_2))$$

This property does not square well with the view of causal strength as difference-making, but it captures a plausible principle for averaging causal strength judgments.

Finally, one can investigate measures of **causal contribution** (e.g., Hall, 2004; Kaiserman, 2016; Beckers and Vennekens, 2017). Suppose we ask what is the stronger cause of a car accident (E): drunk driving ($\mathrm{C}_1$) or bad weather conditions ($\mathrm{C}_2$)? One may answer that $\mathrm{C}_1$ is a stronger cause of E than $\mathrm{C}_2$ if and only if $\mathrm{C}_1$ makes E more expected than $\mathrm{C}_2$. In other words, a cause of an effect is stronger than another cause if it has a higher likelihood of producing the effect. This property, called Effect Production, is appealing in contexts where we want to attribute the occurrence of an event to one of its causes.

**Effect Production** For $\mathrm{C}_1, \mathrm{C}_2 \in \mathcal{C}$ and $\mathrm{E} \in \mathcal{E}$,

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

Cases where $C$ is known with certainty suggest a further adequacy constraint. If two events are logically equivalent given C, it makes sense to treat them the same with respect to the causal strength that C has for them. After all, knowing that C has occurred, we cannot distinguish between them any more. Formally:

**Conditional Equivalence** Assume that $E_1$ and $E_2$ are logically equivalent given C. Then $\eta(C, E_1) = \eta(C, E_2)$.

It is easy to show that the Conditional Equivalence property characterizes the Pearl-Halpern measure $\eta_{ph}$:

**Theorem 6 (Representation Theorem for $\eta_{ph}$)** *All measures of causal strength that satisfy Generalized Difference-Making with contrast class $C' = \Omega_C$ and Conditional Equivalence are ordinally equivalent to*

$$\eta_{ph}(C, E) = p^*(E|C)$$

The Pearl-Halpern measure $\eta_{ph}$ has been defended for measuring actual causal power (Halpern and Pearl, 2005a,b). It is also used in proposals for determining causal contributions among several causes of an event (Kaiserman, 2016, 2017). 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).

The Pearl-Halpern measure $\eta_{ph}(C, E) = p^*(E|C) \approx 1$ 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 $\eta_d(C, E)$, however, disagrees (and so do other contrastive measures that compare C and $\neg$C): due to the presence of the backup assassin, poisoning the wine barely made a difference to the king's death, and $\eta_d(C, E) \approx 0$. The two groups of measures also diverge in cases where an action produces an effect, but by doing so, preempts an even stronger cause.

Here is a line of argument for preferring the Pearl-Halpern explication $\eta_{ph}$: contrastive causal strength measures such as $\eta_d$ judge poison in the wine as a weak cause when a backup cause is present (the second assassin). But there is a sense in which poisoning is *always* a strong cause of death. Routine vaccinations are similar examples. We would not say that a vaccine is causally ineffective just because the overall risk of contracting the disease is low. By

relativizing our explications to particular contexts, we seem to have lost an important aspect of strong causes: the capacity to secure an effect in a large variety of circumstances.

However, we can reconstruct this intuition in the proposed account, too. First, not each cause that secures the effect is universally strong. An umbrella is generally sufficient for protecting the person carrying it from getting wet. In the context of a desert climate, however, we would hesitate to identify umbrellas as strong causes of staying dry. Similarly, vaccinations can be more effective in some contexts and less effective in others. Think of a yellow fever vaccination, for example. If somebody travels to a region where yellow fever is endemic and gets vaccinated beforehand, contrastive causal strength measures deliver the right result: vaccination is highly efficacious ($p^*(\text{Disease Contraction} \mid \text{Vaccination, Exposure}) \gg p^*(\text{Disease Contraction} \mid \text{No Vaccination, Exposure})$). In No Exposure contexts, however, vaccination makes almost no difference to the risk of contracting yellow fever. The reason that most people don't seek yellow fever vaccination is that for them, the relevant context is No Exposure. Causal sufficiency, which is the concept explicated by the Pearl-Halpern measure $\eta_{ph}$, is different from causal strength. Crucially, it does not take into account that we only intervene on a cause if we believe the benefits to be substantial.[10]

In general, such examples show that causal strength has a plurality of senses supported by our intuitions, not all of which are explicated by $\eta_d$. But exhausting these senses was never the goal of this paper. Rather, I explicated causal strength as difference-making ("how would $E$ change if I intervened on $C$?"). Within that perspective, the arguments for $\eta_d$ remain compelling—at least to the degree that $\eta_d$ is an excellent default measure and that the choice of other measures requires special justification. Tables 2 and 3 give an overview over which measure satisfies which adequacy condition, and how the representation theorems relate to each other.

[10]However, if immunity to the yellow fever virus rather than contraction risk is the desired effect, then the vaccination turns out to be a very strong cause independent of the choice of context (Exposure vs. No Exposure).

| | Property | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Measure | GDM($\Omega_C$) | GDM($\neg$C) | CPS | WCPS | CE | EP | SE | MUL | NDIE | NDIEP | CC |
| Pearl-Halpern ($\eta_{ph}$) | yes | (yes) | no | yes | yes | yes | no | no | no | no | no |
| Probability Raise ($\eta_{pr}$) | yes | no | yes | yes | no | yes | yes | no | no | no | no |
| Difference ($\eta_d$) | no | yes | yes | yes | no | no | yes | yes | no | no | no |
| Probability Ratio ($\eta_r$, $\eta_{r'}$) | no | yes | no | no | no | no | no | no | yes | yes | no |
| Cheng/Good I ($\eta_{cg}$) | no | yes | yes | yes | no | no | no | no | no | yes | no |
| Cheng/Good II ($\eta_c$, $\eta_g$) | no | yes | no | no | no | no | no | no | no | no | no |
| Conjunctive Closure ($\eta_{cc}$) | no | yes | no | no | no | no | no | no | no | no | yes |

Table 2: A classification of different measures of causal strength according to the adequacy conditions they satisfy. GDM = Generalized Difference-Making (in both versions), CPS = Causation-Prevention Symmetry, WCPS = Weak Causation-Prevention Symmetry, CE = Conditional Equivalence, EP = Effect Production, SE = Separability of Effects, MUL = Multiplicativity, NDIE = No Dilution for Irrelevant Effects, NDIEP = No Dilution for Irrelevant Effects (Prevention), CC = Conjunctive Closure.

| **Contrast Class** $C' = \Omega_C$ | | |
|---|---|---|
| Separability of Effects | $\eta_{pr}$ | Theorem 1 |
| Conditional Equivalence | $\eta_{ph}$ | Theorem 6 |
| **Contrast Class** $C' = \neg C$ | | |
| Separability of Effects | $\eta_d$ | Theorem 1 |
| Multiplicativity | $\eta_d$ | Theorem 2 |
| No Dilution | $\eta_r$, $\eta_{r'}$ | Theorem 3 |
| No Dilution (Prevention) + WCPS | $\eta_{cg}$ | Theorem 4 |
| Conjunctive Closure | $\eta_{cc}$ | Theorem 5 |

Table 3: An overview of the measures with the adequacy conditions that characterize the representation theorems.

# 5 Conclusion

This paper provides axiomatic foundations for a probabilistic theory of causal strength within the causal Bayes nets framework. Synthesizing ideas from the probability-raising and the interventionist view of causation, I have proposed to formalize causal strength as a function of the probability difference that interventions on a cause C make to the effect E.

I have characterized various measures of causal strength in terms of representation theorems, derived from a set of 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. By doing so, this paper creates a methodological bridge to other projects in formal

epistemology, such as explications of degree of confirmation, coherence, and explanatory power.

On the basis of these representation results, I have put forward arguments for using $\eta_d(C, E) = p^*(\mathrm{E}|\mathrm{C}) - p^*(\mathrm{E}|\neg\mathrm{C})$ as a default measure of causal strength. Indeed, also Holland (1986) and Pearl (2001) build their discussion of (path-specific) causal effects on $\eta_d$ as underlying the baseline measure. However, they do not provide a philosophical defense of their choice—a gap which this paper closes. The theoretical analysis also agrees with practice- and decision-oriented arguments for $\eta_d$, as pointed out in the previous sections.

What remains to do? First, I aim at linking this framework to questions about the magnitude of a causal effect, such as the difference of group means (e.g., Cohen's $d$ or Glass's $\Delta$). $\eta_d$, for once, might be extended naturally into this direction.

Second, this work can be connected to information-theoretic approaches to causal specificity (Weber, 2006; Waters, 2007; 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?

Third, 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)?

Fourth, I would like to spell out how this model connects to research on actual causation and the significance of (statistical) normality and (prescriptive) norms in causal reasoning (Knobe and Fraser, 2008; Hitchcock and Knobe, 2009; Halpern and Hitchcock, 2015; Kominsky et al., 2015; Icard et al., 2017).

Fifth, this research has implications for probabilistic explications of explanatory power (McGrew, 2003; Schupbach and Sprenger, 2011; Crupi and Tentori, 2012, e.g.,). In spite of the tight conceptual connection between explanatory power and causal strength (Eva and Stern, 2017), the measures cannot be easily related to each other. A possible reason is that explanatory power focuses—

unlike most investigated measures of causal strength—on how C reduces (subjectively perceived) surprise in E. That is, it compares $p^*(\mathrm{E}|\mathrm{C})$ and $p^*(\mathrm{E})$. Exploring this relationship is another challenge for further research.

These are all open and exciting questions, and it is not difficult to come up with others. I hope that the results presented in this paper are promising enough to motivate further pursuit of an axiomatic theory of causal strength.

# References


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

Beckers, S. and Vennekens, J. (2017). A Principled Approach to Defining Actual Causation. *Synthese*.

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.

Broadbent, A. (2013). *Philosophy of Epidemiology*. Palgrave Macmillan, Basingstroke.

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

Carnap, R. (1950). *Logical Foundations of Probability*. University of Chicago Press, Chicago.

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*, 104:367–405.

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.

Crupi, V. (2013). Confirmation. In Zalta, E., editor, *The Stanford Encyclopedia of Philosophy*.

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

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

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

Eva, B. and Stern, R. (2017). Causal Explanatory Power. *The British Journal for the Philosophy of Science*.

Fitelson, B. (1999). The plurality of Bayesian measures of confirmation and the problem of measure sensitivity. *Philosophy of Science*, 66:S362–S378.

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*, 11:305–318.

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

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

Hall, N. (2004). The Intrinsic Character of Causation. In Collins, J., Hall, N., and Paul, L., editors, *Oxford Studies in Metaphysics*, pages 225–276. Oxford University Press, Oxford.

Halpern, J. (2016). Sufficient Conditions for Causality to Be Transitive. *Philosophy of Science*, 83:213–226.

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*, 56:843–887.

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

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

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

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.

Icard, T. F., Kominsky, J. F., and Knobe, J. (2017). Normality and Actual Causal Strength. *Cognition*, 161:80–93.

Kaiserman, A. (2016). Causal Contribution. *Proceedings of the Aristotelian Society*, 116:387–394.

Kaiserman, A. (2017). Partial Liability. *Legal Theory*, 23:1–26.

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

Kominsky, J. F., Phillips, J., Gerstenberg, T., Lagnado, D., and Knobe, J. (2015). Causal Superseding. *Cognition*, 137:196–209.

Korb, K. B., Nyberg, E. P., and Hope, L. (2009). Information-Theoretic Causal Power. In Emmert-Streib, F. and Dehmer, M., editors, *Information Theory and Statistical Learning*, pages 231–265. Springer, Berlin.

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. (1986). Postscript to "Causation". In *Philosophical Papers*, volume 2, pages 172–213. Cambridge University Press.

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

McGrew, T. (2003). Confirmation, Heuristics, and Explanatory Reasoning. *British Journal for the Philosophy of Science*, 54:553–567.

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

Meijs, W. (2005). *Probabilistic measures of coherence*. PhD thesis, Erasmus University Rotterdam.

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.

Poole, C. (2010). On the origin of risk relativism. *Epidemiology*, 21:3–9.

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 Law Review*, 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*, 70:41–55.

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

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

Spirtes, P., Glymour, C. N., and Scheines, R. (2000). *Causation, prediction, and search*. Springer, New York, 2nd edition.

Sprenger, J. and Stegenga, J. (2017). 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.

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

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. (2012). Causation and Manipulability. In *The Stanford Encyclopedia of Philosophy*.