Causal Conditionals, Tendency Causal Claims and Statistical Relevance

Michał Sikorski,* Noah van Dongen,† and Jan Sprenger‡

December 13, 2019

Abstract

Indicative conditionals and tendency causal claims are closely related to each other (e.g., Frosch and Byrne 2012), but despite these connections, they are usually studied separately. A unifiying framework could consist in their dependence on probabilistic factors such as statistical relevance, but theoretical research along these lines (e.g., Eells 1991; Douven 2008, 2016) needs to be strengthened by more empirical results. This paper closes that gap and presents empirical results on how judgments on tendency causal claims and indicative conditionals are driven by probabilistic factors, and how these factors (in particular statistical relevance) differ in their predictive power for both causal and conditional claims.

Keywords— indicative conditionals; causal conditionals; tendency causal claims; Adams' Thesis; causal strength; statistical relevance

---

*Center for Logic, Language and Cognition (LLC), Department of Philosophy and Educational Science, Università degli Studi di Torino, Via Sant'Ottavio 20, 10124 Torino, Italy. Email: michal.sikorski@unito.it.

†Center for Logic, Language and Cognition (LLC), Department of Philosophy and Educational Science, Università degli Studi di Torino, Via Sant'Ottavio 20, 10124 Torino, Italy. Email: noah.vandongen@unito.it.

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

1 Introduction and Theoretical Background

Indicative conditionals—that is, conditionals that do not involve the auxiliary verb "would" and do not imply falsity of the antecedent—are important linguistic structures. We use them to predict and explain events, to formulate instructions, and to describe causal relationships. For example, we can describe a causal tendency using explicit causal wording:

(1a) A lot of rain causes the ground to be waterlogged.

or by means of an indicative conditional

(1b) If it rains a lot, then the ground will be waterlogged.

This paper presents an empirical study of **causal conditionals**: that is, indicative conditionals that express a causal relationship. From now on, we will refer to them simply as "conditionals". This focus excludes, for example, Dutchman conditionals and purely inferential conditionals.

Studying the connection between conditionals and causal claims is interesting for many reasons. If there is a link, we could develop a unified semantics for both types of expressions. Indeed, empirical evidence shows that indicative conditionals such as (1b) are often paraphrased by causal claims such as (1a) (e.g., Frosch and Byrne 2012). However, in order to formulate a plausible hypothesis between both types of sentences, we have to distinguish between the causal claims which are related to conditionals and those that are not.

A suitable distinction has been proposed by Christopher Hitchcock (2001, 219–220). According to the author, we can distinguish two kinds of causal claims, which have related but different meanings. The verb "to cause" used in actual causal claims is a success verb: it implies that the events described by its arguments took place. Because of this feature, actual causal claims are used for explaining occurrences of particular events, for example:

(2) James Dean's recklessness caused his accident.

There does not seem to be a systematic connection between the truth of an actual causal claim and the truth of the corresponding (indicative) conditional. The success aspect of an actual causal claim—that both cause and effect occurred—cannot be expressed adequately by means of an indicative conditional, implying that the truth conditions for both types of sentences should be different, too.

Therefore we focus on **tendency causal claims**, which describe the general tendency of a cause to bring about an effect, without implying the actual existence of the causal relata. An example of such a claim is (1a) or:

(3a) Pressing the red button causes the fire alarm to go off.

The corresponding conditional is (example taken from Declerck and Reed 2012):

(3b) If you press the red button, the fire alarm goes off.

True conditionals seem to correspond systematically to true tendency causal claims. For instance, if (1a) or (3a) are true, the corresponding conditionals (1b) and (3b) should also be true. Examining this relationship in detail will be a major goal of our paper. From now on, we use the term "causal claims" as a shortcut for tendency causal claims.

In the above cases, the link between the cause and effect was quite strong. Some tendency causal claims express a weaker relationship: the cause raises the probability of the effect, but the effect may not be likely even in the presence of the cause. An example of such case is:

(4a) Smoking causes cancer.

We would classify this sentence as true, but the corresponding conditional:

(4b) If one smokes, one will get cancer.

seems false.

What is the difference between a true tendency causal claim, which supports the corresponding conditional, like (1a) or (3a), and those that do not, like (4a)? A plausible explanation is the difference in **strength of the causal link** between antecedent and consequent. Causal strength is typically interpreted as the degree to which the cause C brings about the effect E. While pressing the red button triggers the fire alarm for sure, smoking increases your risk of developing cancer—but not to a level where it is more likely than not. Intuitively, fire alarm is stronger determined by pressing the button than presence of lung cancer by smoking.

To express the concept of causal strength and causal relevance, probability enters the game. Probabilistic theories of causation classify C as a cause of E if and only if C raises the probability of E in all/some/the relevant contexts (Suppes 1970; Cartwright 1979; Eells 1991; Hitchcock 2016). The probability-raising intuition is also present in interventionist accounts of causation (Pearl 2000; Spirtes, Glymour and Scheines 2000). All this suggests that **statistical relevance between cause and effect** could be a good predictor of whether causal claims are classified as true or false. In particular, a causal claim is more likely to be classified as true the stronger the statistical association is (for discussions of probabilistic measures of causal strength, see Eells 1991; Cheng 1997; Pearl 2001; Fitelson and Hitchcock 2011; Sprenger 2018).

The second possible explanation of the difference between causal claims that support the corresponding conditional and those that do not is the different role of **conditional probability**. Suppositional theories of conditionals (e.g., Ramsey 1926; Adams 1975; Edgington 1995) claim that "If C, then E" *expresses* the subjective degree of belief in E given C. For tendency causal claims, no comparable thesis has been advanced. Conditional probability is also frequently used for predicting judgments on causal conditionals. For example, Ernest Adams (1965, 1975) explicates the acceptability of a conditional by the formula

$$Acc(C \to E) = p(E|C). \qquad \text{(Adams' Thesis)}$$

Similarly, Stalnaker (1968, 1975) develops a propositional semantics of conditionals where they have a definite truth value and their probability corresponds to that very same conditional probability

$$p(C \to E) = p(E|C) \qquad \text{(Stalnaker's Thesis)}$$

3

suggesting that the subjective degree of belief $p(E|C)$ could be an adequate predictor of how causal conditionals are classified (for well-known theoretical problems with Stalnaker's Thesis, see Lewis 1976). Both Adams' and Stalnaker's Thesis are widely supported by patterns observed in natural language reasoning (Adams 1975; Evans et al. 2007; Over et al. 2007; Egré and Cozic 2011; Douven and Verbrugge 2013; Over 2016). On the other hand, recent papers point out that the connection between conditionals and conditional probability crumbles when the antecedent is not (statistically, causally, or otherwise) relevant for the consequent (e.g., Douven and Verbrugge 2010; Skovgaard-Olsen, Singmann and Klauer 2016). Alternative accounts such as the Evidential Support Theory (EST, Douven 2008, 2016) therefore demand not only that the conditional probability of E given C be high, but also that C raise the probability of E. In other words, C needs to provide evidential support for E (see also Krzyzanowska 2015; Krzyżanowska, Collins and Hahn 2017; Crupi and Iacona 2019).[1]

In this paper, we study the commonalities and differences between classifying causal and conditional claims as true or false from an empirical point of view, especially with respect to the role of probability and statistical relevance in predicting these classifications. More precisely, we investigate whether statistical relevance measures affect causal or conditional claims more strongly. There are many experiments devoted both to conditionals and to causal claims (e.g., Frosch and Byrne 2012; Sloman and Lagnado 2015, —see also Douven 2016 for a survey) but, as far as we know, there are no experiments devoted directly to testing the relation between both kinds of expressions. This paper aims to fill this gap and to obtain a comprehensive picture of the factors that predict the classification of causal conditionals. In the next section we formulate the hypotheses that we test in our experiment.

2   The Hypotheses

The baseline idea of our paper is that causal and probabilistic factors predict the judgment of a causal conditional as true or false. While this claim is in agreement with most of the theoretical and empirical literature, it is too vague to be tested experimentally.

Our first hypothesis concerns the logical relationship between causal conditionals and the corresponding tendency causal claims. Two different kinds of relation are possible:

H1.a (Necessity) Causal conditionals are classified as true *only if* the corresponding tendency causal claim is classified as true.

H1.b (Sufficiency) Causal conditionals are classified as true *if* the corresponding tendency causal claim is classified as true.

H1.a seems prima facie plausible since absence of a causal relation has undermined the acceptability of conditionals in previous empirical research (e.g., Skovgaard-Olsen, Singmann and Klauer 2016). On the other hand, in the light of the above examples (e.g., the pair (4a)/(4b)),

---

[1]The evidential support theory can be maintained either for truth conditions or for acceptability conditions; both varieties are to be taken seriously.

we would expect that H1.b fails in empirical investigation. Some causal conditionals are expected to be classified as false although the corresponding tendency causal claim appears true (e.g., "smoking causes cancer"). We operationalize these hypotheses by demanding that of all causal conditionals evaluated as true, only a small percentage of the corresponding tendency causal claims are evaluated as false (H1.a). Similarly, for an overwhelming percentage of all tendency causal claims evaluated as true, the same claim in conditional form needs to be evaluated as true (H1.b). For the respective thresholds we consider a strict interpretation (5 and 95%) and a lenient interpretation (10 and 90%); as we will see later, this difference in interpretation does not affect the conclusions we draw.[2]

Similarly, we would like to test whether statistical relevance measures predict the evaluation of a causal claim as true or false. We focus on measures that depend on $p(E|C)$ and $p(E)$ since these quantities are the easiest to elicit and form a natural class of statistical relevance measures (i.e., they all conform to the Final Probability Incrementality condition, Crupi 2015; Sprenger and Hartmann 2019, ch. 1).[3] Among them, we consider the measures that are most frequently discussed and defended in the vast literature on probabilistic measures of evidential support (e.g., Fitelson 2001; Crupi, Tentori and González 2007):

$$d(C, E) = p(E|C) - p(E) \quad l(C, E) = \log \frac{p(E|C)}{1 - p(E|C)} - \log \frac{p(E)}{1 - p(E)}$$

$$r(C, E) = \log \frac{p(E|C)}{p(E)} \qquad z(C, E) = \begin{cases} \frac{p(E|C) - p(E)}{1 - p(E)} & \text{if } p(E|C) \geq p(E) \\ \frac{p(E|C) - p(E)}{p(E)} & \text{if } p(E|C) < p(E) \end{cases}$$

(SR)

The logarithmic transformation maps r and l to a scale that is adequate for predictor variables in a Generalized Linear Mixed Model (GLMM). All measures can be motivated intuitively: $d(C, E)$ and $r(C, E)$ are natural indicators of how much supposing C increases the probability of E on an additive and a multiplicative scale. The measure $l(C, E)$ corresponds to the ratio between posterior and prior odds on E when C is learned in the meantime (i.e., the Bayes factor). Finally, $z(C, E)$ can be conceptualized as a generalization of logical entailment between C and E to the case of uncertain reasoning. We can now formulate our hypothesis:

H2.a Statistical relevance measures in the class (SR) predict the classification of a tendency causal claim as true or false.

As a criterion for evaluating H2.a, we adopt the statistical significance of including statistical relevance as a predictor variable, plus a reasonable, non-negligible effect size. Effect size is measured by how much variance in the data can be explained by the predictor variables and expressed numerically by the squared correlation coefficient R^2. For an effect size to be

---

[2]It would also be interesting to test a quantitative version of H1.a and H1.b: the probability of classifying the conditional claim as true increases with the perceived strength of the causal tendency.

[3]A measure depending on $p(E|C)$ and $p(E|\neg C)$ like the popular measure $\Delta p = p(E|C) - p(E|\neg C)$ would require the participant to make two conflicting suppositions ("Suppose $C/\neg C$ is the case. How likely do you consider E?"). This is something we would like to avoid in this experiment, and therefore we focus on measures that depend on $p(E|C)$ and $p(E)$.

meaningful, we demand that it exceed the value $R^2 = 0.09$, which is conventionally identified with the lower bound of a medium effect (Cohen 1988).

Then, we ask which of the two Causal Bayesian Networks in Figure 1 is more adequate: the picture where probabilistic factors predict causal claims only via statistical relevance, or the picture where they add predictive value on top of statistical relevance. See Figure 1. In other words, the question is whether we can "compress" the two probabilities into a single variable for the purpose of predicting the classification of causal claims. This leads us to our next hypothesis:

H2.b Statistical relevance measures in the class (SR) predict the classification of a tendency causal claim as true or false almost as well as the statistical relevance taken together with conditional probability and probability of the consequent.

As a criterion for whether statistical relevance is an adequate proxy, we demand that the explained variance does not meaningfully increase over and above that what is already explained by the model that contains just statistical relevance as a predictor variable. Similar to H2.a, the difference should not exceed $R^2 = 0.09$.
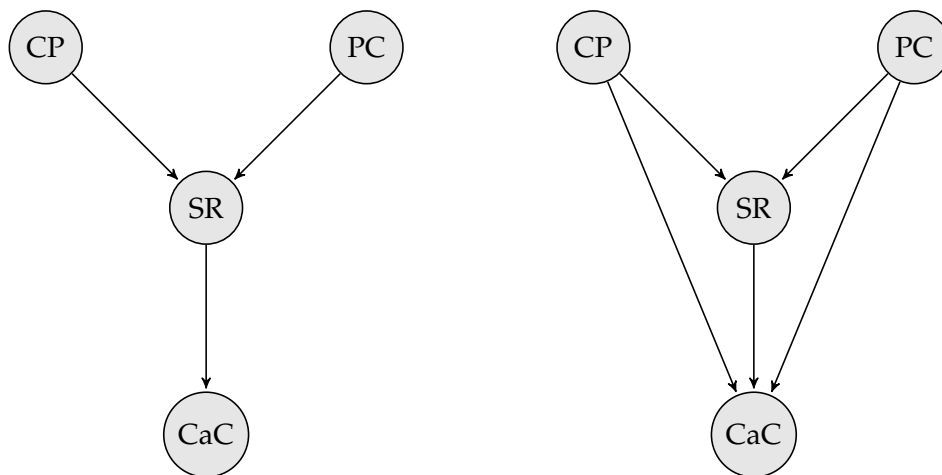

Figure 1: Two possible models for predicting the classification of causal claims, represented as Causal Bayesian Networks. CP = conditional probability, PC = probability of consequent, SR = statistical relevance, CaC = causal claim classification.

Third, while there is substantial debate about the empirical adequacy of Stalnaker's Thesis and Adams' Thesis (Douven and Verbrugge 2013; Douven 2016; Skovgaard-Olsen, Singmann and Klauer 2016), even critics agree on a weak version of those theses: for relevant subclasses of causal conditionals, $p(E|C)$ is moderately correlated with the classification of a conditional as true. We should therefore expect that this probability predicts the classification of a causal conditional at least to some degree.

H3 The conditional probability $p(E|C)$ predicts the classification of causal conditionals of the form "if C, then E" as true or false.

H3 is evaluated on the same basis as the H2.a: adding the conditional probability as a predictor variable needs to be statistically significant, and the effect size as measured by the correlation coefficient must exceed $R^2 = .09$.

Finally, we come to our most interesting hypothesis which posits a specific difference between causal and conditional claims. The intuitive idea is that a conditional is classified as true if and only if two conditions are satisfied: (1) the causal claim is classified as true, in agreement with H1.a; (2) the probability of the consequent, given the antecedent, $p(E|C)$, is "high enough". As explained in the introduction, this hypothesis is supported by research in the tradition of Stalnaker (1968) and Adams (1975) where conditionals are evaluated according to the relevant conditional probability.

H4.a In the class of tendency causal claims classified as true, the conditional probability $p(E|C)$ predicts the evaluation of conditional claims as true.

H4.b In the class of tendency causal claims classified as true, the statistical relevance measures in (SR) predict the evaluation of conditional claims as true.

We expect a positive effect in the first hypothesis and a null effect in the second hypothesis: the effect of statistical relevance on the classification of a conditional is "absorbed" by the evaluation of the causal claim as true. In a Causal Bayesian Network, this would correspond to the claim that the truth of the causal claim screens off statistical relevance from the truth of the conditional claim. In this way, we explain the limitations that probabilistic theories (e.g., Adams' Thesis) experience when causal relevance between antecedent and consequent is suspended. The criteria evaluating for H4.a and H4.b correspond to those for H2.a and H3: statistical significance plus meaningful effect $R^2 > .09$. Our overall picture amounts to modifying Evidential Support Theory as follows: where EST demands, on top of high conditional probability, statistical relevance between C and E for assessing a conditional as true, our account demands in addition the presence of a causal link between C and E. Of course, this need not be in tension with EST since we conjecture that statistical relevance is actually a good predictor of classifying a causal tendency claim as true.

3 Experimental Design and Methods

3.1 Participants

Participants were recruited via Amazon's Mechanical Turk (www.mturk.com). Mechanical Turk directed the participants to the experiment that was run on the Qualtrics platform (www.qualtrics.com). In return for their participation, subjects received a small monetary compensation. Seventy-four native English speakers participated in the experiment. Eighteen participants were excluded because they failed to give the correct response to at least one of the control questions. All participants indicated to have participated seriously. At the end of the experiment, participants were asked an open question about what they thought the experiment was about. None of the participants displayed clear knowledge of the purpose of the

experiment. In total, 56 [39 female, mean age = 40.59 years, s.d. = 11.69 years] participants were included in the analysis.

3.2 Design

We used a within-subjects design were each participant evaluated 19 scenarios. These scenarios were presented in random order on the participants' computer screen. The participants were instructed to answer questions with the requirement that each question needed to be answered to be able to progress to the next item (i.e., forced-choice). The entire experiment was conducted in English.

Control questions were used as a check on participants' attention and participation. Randomly dispersed throughout the experiment, participants had to give a correct answer to several repeats of the *elimination questions* "For quality control, please select answer category five. If you do not select five, the survey will be terminated." In addition, subjects had to rate on a five-point Likert scale how seriously they participated in the experiment (1= "completely unserious", 5 = "completely serious"). We included an open question on the purpose of the experiment, because knowledge of the experiment could influence the behaviour of participants. Our exclusion criteria were: failure to give the answer 5 on one of the elimination questions; rating their seriousness in participation as 3 or lower; or describing the experiment as being about conditionals and causation or something similar.

3.3 Material and Procedure

Participants had to evaluate tendency causal claims and the corresponding conditional claims in hypothetical scenarios, where a certain situation was described (e.g., the effect of prohibiting alcohol consumption on the crime rate). Through several pre-studies, 19 scenarios were selected on comprehensibility from a list of 60. When directed from Mechanical Turk to Qualtrics, participants first received instructions on the experiment. After the instructions they were presented with the scenarios in a randomized sequence. The 19 scenarios consisted of four questions each, eliciting probability judgments as well as dichotomous judgments on causal and conditional claims:

(Unconditional) Probability of Consequent This question elicits the probability of a certain development without making specific assumptions (e.g., "how likely is it that the crime rate will decline in the next five years?").

Conditional Probability of the Consequent This question elicits the probability of the same development under a specific assumption stated in the antecedent (e.g., "how likely is it that the crime rate will decline in the next five years if alcohol consumption is made illegal?").

Causal Claim This question asks the participants to evaluate the truth or falsity of the causal connection between antecedent and consequent (e.g., "Making alcohol consumption illegal will cause the crime rate to decline in the next five years").

Conditional Claim This question asks the participants to evaluate the truth or falsity of the corresponding indicative conditional (e.g. "If alcohol consumption is made illegal, then the crime rate will decline in the next five years.")

The first two questions had to be answered on a visual analog scale of probability percentages from 0% to 100%. The third and fourth question had to be answered with either "true" or "false". We reproduce the entire experimental material in Appendix A and B.

4 Results

4.1 Hypotheses H1.a and H1.b

Combined, the data consisted of 1064 entries; 56 participants responded to 19 scenarios. We evaluated H1.a and H1.b by simply checking the frequency statistics for the relevant categories (see Table 4.1). Of the 531 data points where a causal conditional was classified as true, 25 data points classified the corresponding tendency causal claim as false. This corresponds to a percentage of 4,71% and therefore confirms our hypothesis H1.a that perceived presence of a causal relationship is *necessary* for classifying a causal conditional as true, both for the strict 5% and the lenient 10% threshold. By contrast, Hypothesis H1.b that classifying a causal conditional as true is a *sufficient* condition for classifying the causal conditional as true was not borne out by the data: of 611 data points where the tendency causal claim was evaluated as true, only 506 evaluated the corresponding conditional as true. This percentage of 82,82% is clearly below the thresholds of 90% and 95% necessary to establish sufficiency.

4.2 Hypotheses H2.a and H2.b

A Generalized Linear Mixed Model (GLMM) was used to test hypothesis H2.a and H2.b. We used a logit link function, as the outcome variable for each hypothesis was binary (0 = False, 1 = True). We added participants and scenario number as crossed random effects, because of difference in content between the scenarios, and possible differences in their interpretation between participants. We used the R package *lme4* (Bates et al. 2015) to estimate the GLMM's regression coefficients, variance components, and the amount of variance in the outcome explained by the predictor(s).

All statistical relevance measures predict the classification of tendency causal claims as true or false, thus supporting hypothesis H2.a (see Table 2). For all statistical relevance measures,

Contingency Table		Conditional Claim		Total
		True	False	
Causal Claim	True	506	105	611
	False	25	428	453
Total		531	533	1064

Table 1: Classification of causal and conditional claims as true and false.

the analyses show a positive and meaningful association ($R^2 > 0.09$) with the proclivity of participants to assess the tendency causal claim as true. Specifically, the coefficients indicate the estimated increase in log odds (per unit of the relevance measure) of tendency causal claims being indicated as true versus false.[4] The random-effects for scenario and participant, though non-zero, appear to be minor. Specifically, their coefficients are only slightly larger than their standard errors. Based on the test statistics (z-values) and amount of explained variance (R^2), the association between $z(C, E)$ and classification of tendency causal claims was the strongest (i.e., largest coefficient with respect to its standard error). The weakest association was with the $r(C, E)$ measure.

To test hypothesis H2.b, the change in R^2 is assessed when *probability of the consequent* $p(E)$ and *conditional probability* $p(E|C)$ are added to the models of H2.a. Unfortunately, these predictors cannot be added to the $d(C, E)$ model, because they perfectly define, without transformation, this statistical relevance measure (see the SR equations). In other words, these predictors are incapable of explaining additional variance in peoples tendency to indicate causal claims as true over and above $d(C, E)$. For the other three statistical relevance measures, only $z(C, E)$ is an adequate proxy for the effects of *probability of the consequent* and *conditional probability*. Specifically, when the two probability predictors were added to the models the R^2s increased to about 0.47 for all statistical relevance measures, meaning that only $z(C, E)$ is a good proxy for the probability variables (i.e., R^2 difference is below the 0.09 threshold; $0.47 - 0.40 = 0.07$).

4.3 Hypothesis H3

To test H3, a GLMM similar to the test of H2 was used. Again, we used a logit link function and added participants and scenario number as crossed random effects. The results show a strong and positive association between the conditional probability ascribed to the scenarios and the log odds of the corresponding causal conditional being considered as true. Specifically, with every percentage-point increase in conditional probability these log odds are estimated to increase by 0.07 (an increase of 7 over the full 100 percentage points). Most importantly, the model explains 32% ($R^2 = 0.32$) of the variance in the participants tendency to indicate causal conditionals as either true or false. In short, H3 is supported by the observed data.

4.4 Hypotheses H4.a and H4.b

To test hypotheses H4.a and H4.b, only those data points were used where the tendency causal claim was indicated as true (see Table 4.1). For analyses H4.a. and H4.b. we again used GLMMs to estimate the coefficients and the variance explained by the predictors. Similar to the results for H3, results show a positive association between the conditional probability ascribed to the scenarios and the log odds of corresponding causal conditional being considered as true (see Table 4). Although the amount of variance explained is greatly reduced, from 32%

---

[4]Please note that the statistical relevance measures are not measured on the same scale. Thus the coefficients indicated in the tables cannot be meaningfully compared in the sense that the largest coefficient is the best predictor.

Type of Effect	Variable	Coefficient	Std. Error	z	p
Fixed effects	intercept	-0.15	0.31	-0.48	0.63
	$d(C,E)$	5.85	0.53	11.04	<0.0001
Random effects	Participant	1.14	1.07		
	Scenario	1.26	1.12		
Explained variance: $R^2 = 0.34$. Residual degrees of freedom $= 1060$.					
Fixed effects	intercept	0.26	0.35	0.75	0.46
	$r(C,E)$	1.16	0.14	8.36	<0.0001
Random effects	Participant	1.02	1.01		
	Scenario	1.76	1.33		
Explained variance: $R^2 = 0.22$. Residual degrees of freedom $= 1060$.					
Fixed effects	intercept	-0.04	0.32	-0.12	0.90
	$l(C,E)$	0.84	0.08	10.08	<0.0001
Random effects	Participant	1.10	1.05		
	Scenario	1.33	1.15		
Explained variance: $R^2 = 0.34$. Residual degrees of freedom $= 1060$.					
Fixed effects	intercept	-0.24	0.29	-0.84	0.40
	$z(C,E)$	3.74	0.30	12.42	<0.0001
Random effects	Participant	1.10	1.04		
	Scenario	0.99	0.99		
Explained variance: $R^2 = 0.40$. Residual degrees of freedom $= 1060$.					

Table 2: The Generalized Linear Mixed Model (GLMM) for the dependent variable *Causal Claim* as a function of *Statistical Relevance*, quantified by various measures.

to 12%, it is still considered meaningful ($R^2 > 0.09$) and thus supporting H4.a. None of the statistical relevance measures made a meaningful difference in explaining the variance in the participants' tendency to indicate the causal conditionals as true of false ($R^2 < 0.09$ in all cases, see Table 5), thus supporting our conjecture of a null effect and contradicting hypothesis H4.b.

Type of Effect	Variable	Coefficient	Std. Error	z	p
Fixed effects	intercept	-4.61	0.46	-10.13	<0.0001
	Conditional Probability	0.076	0.006	13.06	<0.0001
Random effects	Participant	1.84	1.36		
	Scenario	0.64	0.80		
Explained variance: $R^2 = 0.34$. Residual degrees of freedom $= 1060$.					

Table 3: The Generalized Linear Mixed Model (GLMM) for the dependent variable *Causal Claim* as a function of *Conditional Probability*.

Type of Effect	Variable	Coefficient	Std. Error	z	p
Fixed effects	intercept	0.89	0.62	-1.14	0.15
	Conditional Probability	0.05	0.008	5.79	<0.0001
Random effects	Participant	3.39	1.84		
	Scenario	0.45	0.67		
Explained variance: $R^2 = 0.13$. Residual degrees of freedom $= 607$.					

Table 4: The Generalized Linear Mixed Model (GLMM) for the dependent variable *Conditional Claim* as a function of *Conditional Probability* when *Causal Claim* = "true".

Type of Effect	Variable	Coefficient	Std. Error	z	p
Fixed effects	intercept	1.89	0.38	4.99	<0.0001
	$d(C,E)$	2.04	0.62	3.28	0.001
Random effects	Participant	2.80	1.67		
	Scenario	0.64	0.80		
Explained variance: $R^2 = 0.04$. Residual degrees of freedom $= 607$.					
Fixed effects	intercept	2.15	0.37	5.82	<0.0001
	$r(C,E)$	0.36	0.16	2.17	0.030
Random effects	Participant	2.72	1.65		
	Scenario	0.65	0.81		
Explained variance: $R^2 = 0.02$. Residual degrees of freedom $= 607$.					
Fixed effects	intercept	1.89	0.38	4.99	<0.0001
	$l(C,E)$	0.37	0.09	4.00	0.0001
Random effects	Participant	2.94	1.71		
	Scenario	0.64	0.80		
Explained variance: $R^2 = 0.07$. Residual degrees of freedom $= 607$.					
Fixed effects	intercept	1.63	0.38	4.31	<0.0001
	$z(C,E)$	1.73	0.41	4.26	<0.0001
Random effects	Participant	2.70	1.64		
	Scenario	0.60	0.78		
Explained variance: $R^2 = 0.07$. Residual degrees of freedom $= 607$.					

Table 5: The Generalized Linear Mixed Model (GLMM) for the dependent variable *Conditional Claim* as a function of *Statistical Relevance* (quantified by various measures) when *Causal Claim* = "true".

5 Evaluation and Discussion

There is a natural mapping between tendency causal claims ("Smoking weed causes dizziness") and causal conditionals in the indicative mood ("if somebody smokes weed, he will feel dizzy"). In the presented study, we tested various hypotheses about the classification of such sentences as true or false, especially with respect to predicting these classifications as a function of statistical relevance. This is a highly relevant research question since the influence of probabilistic factors on causal claims and causal conditionals has been studied extensively, but in separate literatures. Unlike for counterfactuals, we do not yet have a semantic or pragmatic theory connecting both types of claims. We therefore conducted a study where participants classified a given causal claim and the corresponding conditional as true or false, and estimated in addition two probabilistic variables: the conditional probability of the consequent, given the antecedent, and the probability of the consequent simpliciter. Our specific interest was in finding whether these probabilistic features could reliably predict the classification of causal claims/causal conditionals as true or false.

Our informal discussion at the beginning of this paper has suggested that the conditions for classifying a causal conditional as true will be more demanding than the conditions for classifying the corresponding tendency causal claim as true. This claim, expressed in hypotheses H1.a and H1.b, has been supported convincingly by our experimental results. Then, we built on probabilistic theories of causation and causal strength in order to formulate our second pair of hypotheses: classification of a causal claim as true or false can be reliably predicted by measures of statistical relevance (H2.a). More precisely, statistical relevance measures can act as a good proxy for the full probabilistic data in a Generalized Linear Mixed Model (GLMM): predicting the classification of tendency causal claims by statistical relevance explains almost as much of the variance in the data as using all available probabilistic predictors (H2.b).

These hypotheses have been confirmed by our data, but H2.b has to be qualified since the outcome depends on the choice of the statistical relevance measure. Only the z-measure—a generalization of logical entailment to uncertain reasoning—meets our prespecified effect size threshold. This phenomenon is a classical case of *measure sensitivity*: the theoretical and empirical properties of statistical relevance, incremental confirmation or evidential support depend on the particular measure chosen (Fitelson 1999, 2001). We can only speculate why z fares better than l, and clearly better than d and r in our case. One plausible explanation would be that the z-measure normalizes statistical relevance, as measured by d, by *maximal* statistical relevance—that is, raising the probability of E to unity. This means that the same probability difference counts more if the conditional probability is high than when it is low. This feature would match with intuitions that *ceteris paribus*, causes are stronger if they are likely to bring about an effect (e.g., as expressed in regularity theories or necessitarian theories of causation). Apart from its contribution to studying causal and conditional judgments, this paper thus adds to an already existing literature that singles out z as a particularly apt measure of statistical relevance—both on the basis of axiomatic representation theorems and empirical performance (Crupi, Tentori and González 2007; Crupi and Tentori 2013).

For the classification of the causal conditional, we have established that the conditional probability is indeed a reliable predictor, in line with a weak version of Adams' Thesis (H3). The effect size in the GLMM is very remarkable ($R^2 = 0.32$). This relationship remains intact when restricted to the set of causal claims classified as true (H4.a) although the effect size decreases to $R^2 = 0.12$, showing that the classification of the causal claim already predicts a good part of the classification of the causal conditional.

However, statistical relevance is not any more a relevant predictor if the corresponding causal claim is classified as true (H4.b): the relationship between the various measures and the target variable is still statistically significant, but this is to be expected for such a large data set and the effect size is too small to be of theoretical interest ($R^2 < 0.09$ for all statistical relevance measures).

If we extract a theory of how reasoners classify causal conditionals, our results suggest that causal conditionals are supported by sufficiently strong causal claims: (a) the corresponding tendency causal claim needs to be classified as true and (b) the odds for classifying the conditional as true co-vary with the conditional probability $p(E|C)$. Statistical relevance can act as a proxy for the first component (especially when the z-measure is used), but not for the classification of the conditional. Figure 2 gives a schematic representation of the relationship between the different variables in our model, adding a latent variable that expresses the part of the causal classification judgment that cannot be reduced to statistical relevance, or other probabilistic predictors.
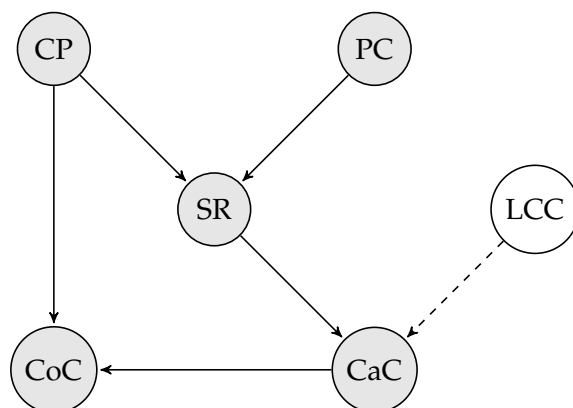

Figure 2: Predictors of the classification of a conditional claim (CoC) as true or false, represented via a Causal Bayesian Network. CP = conditional probability, PC = probability of consequent, SR = statistical relevance, CaC = causal claim classification, LCC = latent causal component.

All in all, this account resembles Igor Douven's Evidential Support Theory (Douven 2008, 2016; Douven and Verbrugge 2012) and more recent developments along these lines (Crupi and Iacona 2019), and it provides empirical support for these research programs. Our account is different in that the explicit consideration of a causal component adds a "semantic" component which is absent in those proposals.

One of the limitations of our study is the exclusion of counterfactual conditionals, whose causal character has been studied extensively in the literature (Lewis 1973a,b; Pearl 2000; Schulz 2017). We conjecture that analogical relations might hold between counterfactual conditionals and counterfactual causal claims. Consider, for example, the sentences "If Ben had gone to the party, he would have failed his exam" and "Attending the party would have caused him to fail his exam". The relationship between such pairs of sentences strikes us as a valuable object for further research (see also Schulz 2011). In addition, we might study whether the results change when we ask when we ask for the *assertability* of conditionals instead of truth (reasons for suspecting invariance are given by Douven and Krzyzanowska 2018), or when we replace a dichotomous choice by a continuous scale. Finally, we might elicit the probability of the effect given the negation of the cause (i.e., $p(E|\neg C)$) and see whether statistical relevance measures that depend on this quantity lead to different qualitative results. All in all, the interface of conditionals, causality and statistical relevance emerges as an important and fruitful area for future research.

References

Adams, Ernest W. (1965). The Logic of Conditionals. *Inquiry 8*, 166–197.

Adams, Ernest W. (1975). *The Logic of Conditionals*. Dordrecht: Reidel.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software 67*, 1–48.

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

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

Cohen, Jacob (1988). *Statistical Power Analysis for the Behavioral Sciences*. Newark, N.J.: Lawrence & Erlbaum.

Crupi, Vincenzo (2015). Confirmation. In Ed Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Retrieved from https://plato.stanford.edu/archives/sum2018/entries/confirmation/.

Crupi, Vincenzo and Andrea Iacona (2019). Three ways of being non-material. Unpublished manuscript.

Crupi, Vincenzo and Katya Tentori (2013). Confirmation as Partial Entailment: A Representation Theorem in Inductive Logic. *Journal of Applied Logic 11*, 364–372.

Crupi, Vincenzo, Katya Tentori, and Michel González (2007). On Bayesian Measures of Evidential Support: Theoretical and Empirical Issues. *Philosophy of Science 74*, 229–252.

Declerck, Renaat and Susan Reed (2012). *Conditionals. A Comprehensive Empirical Analysis*. Mouton de Gruyter.

Douven, Igor (2008). The Evidential Support Theory of Conditionals. *Synthese 164*, 19–44.

Douven, Igor (2016). *The Epistemology of Indicative Conditionals*. Cambridge: Cambridge University Press.

Douven, Igor and Karolina Krzyzanowska (2018). The Semantics-Pragmatics Interface. In Alessandro Capone (ed.), *Further Advances in Pragmatics and Philosophy*, Volume 2. Springer.

Douven, Igor and Sara Verbrugge (2010). The Adams family. *Cognition 117*, 302–318.

Douven, Igor and Sara Verbrugge (2012). Indicatives, concessives, and evidential support. *Thinking & Reasoning 18*, 480–499.

Douven, Igor and Sara Verbrugge (2013). The Probabilities of Conditionals Revisited. *Cognitive Science 37*, 711–730.

Edgington, Dorothy (1995). On Conditionals. *Mind 104*, 235–329.

Eells, Ellery (1991). *Probabilistic Causality*. Cambridge: Cambridge University Press.

Egré, Paul and Mikaël Cozic (2011). If-clauses and probability operators. *Topoi 30*, 17.

Evans, Jonathan, Simon Handley, Helen Neilens, and David Over (2007). Thinking about conditionals: A study of individual differences. *Memory & Cognition 35*, 1772–1784.

Fitelson, Branden (1999). The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity. *Philosophy of Science 66*, S362–S378.

Fitelson, Branden (2001). *Studies in Bayesian Confirmation Theory*. Ph.D. thesis, University of Wisconsin–Madison.

Fitelson, Branden and Christopher Hitchcock (2011). Probabilistic Measures of Causal Strength. In Phyllis McKay Illari, Federica Russo, and Jon Williamson (eds.), *Causality in the Sciences*, pp. 600–627. Oxford: Oxford University Press.

Frosch, Caren A. and Ruth M.J. Byrne (2012). Causal conditionals and counterfactuals. *Acta Psychologica 141*, 54 – 66.

Hitchcock, Christopher (2001). Causal generalizations and good advice. *The Monist 84*, 218–241.

Hitchcock, Christopher (2016). Probabilistic Causation. In Alan Hajek and Christopher Hitchcock (eds.), *Handbook of the Philosophy of Probability*. Oxford: Oxford University Press.

Krzyzanowska, Karolina (2015). *Between "If" and "Then". Towards an empirically informed philosophy of conditionals.* PhD thesis, University of Groningen.

Krzyżanowska, Karolina, Peter J. Collins, and Ulrike Hahn (2017). Between a conditional's antecedent and its consequent: Discourse coherence vs. probabilistic relevance. *Cognition 164*, 199–205.

Lewis, David (1973a). Causation. *Journal of Philosophy 70*, 556–567.

Lewis, David (1973b). *Counterfactuals*. Oxford: Blackwell.

Lewis, David (1976). Probabilities of conditionals and conditional probabilities. *Philosophical Review 85*, 297–315.

Over, David (2016). Causation and the probability of causal conditionals. In Michael Waldmann (ed.), *Oxford Handbook of Causal Reasoning*. Oxford: Oxford University Press.

Over, David, Constantinos Hadjichristidis, Jonathan Evans, Simon Handley, and Steven Sloman (2007). The probability of causal conditionals. *Cognitive Psychology 54*, 62–97.

Pearl, Judea (2000). *Causality*. Cambridge: Cambridge University Press.

Pearl, Judea (2001). Direct and Indirect Effects. In Jack Breese and Daphne Koller (eds.), *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, pp. 411–420.

Ramsey, Frank P. (1926). Truth and Probability. In D H Mellor (ed.), *Philosophical Papers*, pp. 52–94. Cambridge: Cambridge University Press.

Schulz, Katrin (2011). "If You'd Wiggled A, Then B Would've Changed". *Synthese 179*, 239–251.

Schulz, Moritz (2017). *Counterfactuals and Probability*. Oxford: Oxford University Press.

Skovgaard-Olsen, Niels, Henrik Singmann, and Karl Christoph Klauer (2016). The relevance effect and conditionals. *Cognition 150*, 26 – 36.

Sloman, Steven A. and David Lagnado (2015). Causality in Thought. *Annual Review of Psychology 66*, 223–247.

Spirtes, Peter, Clark Glymour, and Richard Scheines (2000). *Causation, Prediction, and Search* (2nd ed.). New York: Springer.

Sprenger, Jan (2018). Foundations for a Probabilistic Theory of Causal Strength. *Philosophical Review 127*, 371–398.

Sprenger, Jan and Stephan Hartmann (2019). *Bayesian Philosophy of Science*. Oxford: Oxford University Press.

Stalnaker, Robert (1968). A Theory of Conditionals. In *Studies in Logical Theory: American Philosophical Quarterly Monograph Series, No. 2*. Oxford: Blackwell.

Stalnaker, Robert (1975). Indicative Conditionals. *Philosophia 5*, 269–286.

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

Appendix A. Instructions for the Experiment

"**Thank you for participating in our study.**

As a Mechanical Turk worker, you should at least have an approval rating of 95% if you want to be reimbursed for your participation.

Please read the questions carefully and answer each with a probability score (between **0%** and **100%**) or truth value ('**True**' or '**False**')

If you have any questions or comments, please leave them at the end of the survey.

You start the survey by pressing the double arrow at the bottom-right corner."

Appendix B. List of Scenarios and Questions

1a. John is a middle-aged man, how likely is it that he will be healthy?

1b. John is a middle-aged man, how likely is it that he will be healthy if he exercises daily?

1c. Daily exercising causes John to be healthy.

1d. If John exercises daily, then he will be healthy.

2a. How likely is it that a random person is skinny?

2b. How likely is it that a random person is skinny if his/her daily food intake is 4 apples and 3 cucumber sandwiches?

2c. Eating only 4 apples and 3 cucumber sandwiches a day causes people to become skinny.

2d. If people only eat 4 apples and 3 cucumber sandwiches a day, then they will become skinny.

3a. How likely is it that a random person will catch the flu?

3b. How likely is it that a random person will catch the flu if two-thirds of his/her co-workers already have it?

3c. Having people around oneself with the flu causes one to catch the flu.

3d. If people around oneself have the flu, then one will catch the flu.

4a. How likely is it that a random person has more than 10 friends?

4b. How likely is it that a random person has more than 10 friends if he/she uses MDMA at parties?

4c. Using MDMA at parties causes people to have more than 10 friends.

4d. If people use MDMA at parties, then they will have more than 10 friends.

5a. How likely is it that the crime rate will decline in the next 5 years?

5b. How likely is it that the crime rate will decline in the next 5 years if drugs (xtc, cocaine, weed) are legalized?

5c. Legalizing drugs (xtc, cocaine, and weed) causes the crime rate to decline over the next 5 years.

5d. If drugs (xtc, cocaine, and weed) are legalized, then the crime rate will decline over the next 5 years.

6a. How likely is it that the crime rate will decline in the next 5 years?

6b. How likely is it that the crime rate will decline in the next 5 years if alcohol consumption is made illegal?

6c. Making alcohol consumption illegal causes the crime rate to decline over the next 5 years.

6d. If alcohol consumption is made illegal, then the crime rate will decline over the next 5 years.

7a. How likely is it that the national birth rate (babies born per capita per year) will decline in the next 10 years?

7b. How likely is it that the national birth rate (babies born per capita per year) will decline in the next 10 years if contraception is mandatory until the age of 21?

7c. Making contraception mandatory until age 21 causes the national birth rate to decline.

7d. If contraception is made mandatory, then the national birth rate will decline.

8a. How likely is it that the national birth rate (babies born per capita per year) will decline in the next 10 years?

8b. How likely is it that the national birth rate (babies born per capita per year) will decline in the next 10 years if all men start wearing white socks in sandals?

8c. All men wearing white socks in sandals causes the national birth rate to decline.

8d. If all men start wearing white socks in sandals, then the national birth rate will decline.

9a. How likely is it that the national birth rate (babies born per capita per year) will decline in the next 10 years?

9b. How likely is it that the national birth rate (babies born per capita per year) will decline in the next 10 years if education is mandatory until age 21?

9c. Making education mandatory until age 21 causes the national birth rate to decline.

9d. If education is mandatory until age 21, then the national birth rate will decline.

10a. How likely is it that a random non-fiction book will be an international best seller (3000 copies sold in the first week)?

10b. How likely is it that a random non-fiction novel will be a best seller (3000 copies sold in the first week) if the author is Stephen Hawking?

10c. Having Stephen Hawking as the author causes a book to be a best seller.

10d. If Stephen Hawking is the author, then a book will be a best seller.

11a. How likely is it that a random non-fiction book will be an international best seller (3000 copies sold in the first week)?

11b. How likely is it that a random non-fiction novel will be a best seller (3000 copies sold in the first week) if the author is Sarah Palin?

11c. Having Sarah Palin as the author causes a book to be a best seller.

11d. If Sarah Palin is the author, then a book will be a best seller.

12a. John is a man in the late sixties, he eats fast food every day and does not exercise. How likely it is that he will develop a cancer?

12b. John is a man in the late sixties, he eats fast food every day and does not exercise. How likely it is that he will develop a cancer if he smokes a pack of cigarettes a day since he was a teenager?

12c. John's smoking will cause him to develop a cancer.

12d. If John smokes, he will develop a cancer.

13a. How likely it is that a random person will develop a cancer?

13b. How likely it is that a random person, exposed to a high dose of gamma radiation, will develop a cancer?

13c. Exposure to a high dose of gamma radiation causes cancer.

13d. If one is exposed to a high dose of gamma radiation, he will develop a cancer.

14a. How likely it is that a random child will be tall?

14b. How likely it is that a random child will be tall, given that she or he is drinking a lot of milk every day?

14c. Drinking a lot of milk everyday causes one to be tall.

14d. If a child drinks a lot of milk every day, he or she will be tall.

15a. How likely it is that a random child will be tall?

15b. How likely it is that a random child will be tall, if she or he is exercising every day?

15c. Exercising causes one to be tall.

15d. If one exercises every day, he or she will be tall.

16a. How likely it is that a random person will become a diabetic?

16b. How likely it is that a random person who eats three apples a day will become a diabetic?

16c. Eating three apples a day cause diabetics.

16d. If one eats three apples a day, he will become a diabetic.

17a. How likely it is that a random person will have all natural teeth at the age of sixty?

17b. How likely it is that a random person will have all natural teeth at the age of sixty, if he or she brushes one's teeth after every meal?

17c. Brushing one's teeth after every meal will cause one to have all natural teeth at the age of sixty.

17d. If one brushes one's teeth after every meal, she or he will have all natural teeth at the age of sixty.

18a. How likely it is that a random person will have all natural teeth at the age of sixty?

18b. How likely it is that a random person will have all natural teeth at the age of sixty, if he or she brushes one's teeth after every meal and visits a dentist once a month?

18c. Brushing one's teeth after every meal and visiting a dentist once a month will cause one to have all natural teeth at the age of sixty.

18d. If one brushes one's teeth after every meal and visits a dentist once a month, she or he will have all natural teeth at the age of sixty.

19a. What is the probability that the approval of the government will decrease in the next few months?

19b. What is the probability that the approval of the government will decrease in the next few months if the majority party proposes legislation that bans alcohol?

19c. A legislation proposal which bans alcohol, made by the majority party, will cause a decrease in approval of the government in the next few months.

19d. If the majority party proposes legislation that bans alcohol, the approval of the government will decrease in the next few months.