On causal claims, contingencies, and inference: How causal terminology affects what we think about the strength of causal links

Simon Stephan (sstepha1@gwdg.de)
Michael R. Waldmann (michael.waldmann@bio.uni-goettingen.de)
Department of Psychology, University of Göttingen,
Gosslerstr. 14, 37073 Göttingen, Germany

Abstract

The communicative goal behind a causal claim like "Smoking causes heart attacks" is to inform recipients about the existence of a causal link between the factors mentioned in the proposition. Different terminologies can be used to accomplish this goal. Sometimes people use formulations of the form "*C* causes *E*", like in the tobacco warning above, and sometimes they use other formulations, such as modal propositions like "*C* can cause / lead to *E*.", or statements like "*C* increases the risk of *E*.". We investigate the hypothesis that different causal structure claims, by means of different terminologies, not only communicate the existence of a causal link but also implicitly elicit intuitions about that link's strength. Experiment 1 revealed that claims like "*C* causes *E*" imply a stronger link than, for example, modal formulations like "*C* can cause *E*". Experiment 2 tested implications of this finding for research on causal structure learning.

Keywords: causality; causal claims; causal reasoning; causal strength; causal structure; causal talk

Smoking is causally related to lung cancer, heart attacks, and many other diseases. To alert customers to the adverse effects that come a long with lighting up, tobacco companies selling their products in the European Union are obliged to print warning labels on their cigarette packages stating the causal relationship between smoking and at least one of its negative consequences. Consider the following three examples of warning statements that can be found on cigarette packs sold in the EU: (1) "Smoking causes heart attacks.", (2) "Smoking increases the risk of blindness.", and (3) "Smoking can kill your unborn child.".

All these statements represent causal claims whose goal it is to communicate the existence of a causal link between smoking and a particular negative outcome. Beyond the communication of the existence of a causal link, however, these different types of causal claims, by means of their different terminologies, also seem to convey different impressions about the causal strengths of the respective links. Compare the claim that (1) "Smoking causes heart attacks" with the claim that (2) "Smoking increases the risk of blindness.". Would someone who has no knowledge about the statistical association between smoking and heart attacks and between smoking and blindness assume the two causal relationships to be equally strong, or would she assume one to be the stronger than the other? Which of the warnings could be more successful in preventing people from smoking? Intuitively, the first claim seems to convey the impression of a stronger causal link than the second claim.

In the present paper, we investigate which intuitions about the strengths of causal relationships are elicited by causal statements that rely on different causal terminologies to communicate the existence of a causal link, like the tobacco warnings introduced above. Several lines of research have investigated how reasoners translate verbal expressions of uncertainty like "probably", "rarely", "perhaps", or "frequently leads to" into numerical values (see, e.g., Meder & Mayrhofer, 2017; Teigen & Brun, 2003). Interestingly, the causal-strength intuitions that are elicited by different causal phrases whose primary aim it is to communicate the existence of a causal link have so far not been investigated. Using an interactive experimental paradigm, we show that subjects who read different types of such causal-structure claims indeed tend to construct different underlying contingencies. Demonstrating this effect is not the only goal we pursue. A second goal is to elucidate and test the implications that this effect has for a central line of research on causal reasoning – research on *elemental causal structure induction*. Elemental causal structure induction refers to the process in which a reasoner learns about a causal link between a single potential cause and effect factor based on observed data. This process involves two aspects: (1) a decision whether a causal link does or does not exist and (2) how strong a potential link is. Previous research (Griffiths & Tenenbaum, 2005) suggested that reasoners predominantly focus on and report the structural decision as to whether a link does or does not exist, even if the test question intends to probe causal strength. In the present paper we show that participants' seem to be much more sensitive to causal strength, however, as we found that their answers to causal structure queries are influenced by causal-strength impressions elicited by different terminologies used in these questions.

Causal claims and implied causal strengths

In everyday life, exact statistical information about causal relationships is often not available and in many situations it is also not important. Often, a rather coarse representation of causal relationships is sufficient to make rational decisions. For example, for a doctor to know that a patient's symptom requires her immediate attention, it will be enough for her to know that the symptom is related to a disease that "in most cases", "frequently", or "often" causes people to die if untreated. In another case in which she knows that a disease only "very rarely" or "almost never" leads to death, she might decide to prioritize another condition of her patient. The two examples illustrate that qualitative notions about the strength of a causal relationship can be conveyed by means of what has been called "verbal expressions of uncertainty" or "ver-

bal probabilities" (cf. Teigen & Brun, 2003). Verbal probabilities have been investigated in various empirical studies, and it has been shown that reasoners are remarkably consistent at translating different verbal expressions of uncertainty into exact numerical values (see Teigen & Brun, 2003, for a review), and that they can, for instance, use purely verbal information to make diagnostic causal inferences (Meder & Mayrhofer, 2017) that are as accurate as those made by reasoners who received exact numerical information.

The pragmatic goal behind causal claims involving verbal probabilities is the communication of information about the *strength* of causal links. Such verbal strength claims can be contrasted with causal claims like the three tobacco warnings presented in the introduction, whose primary goal it is to communicate the *existence* of a particular causal link. The three examples also illustrate that this goal may be achieved by means of different types of causal terminology like "causes", "can cause", "increases the risk of", and so on. The hypothesis of the present paper is that such "causal structure claims", due to the different types of causal terminology they involve, still *implicitly* convey information about the strength of the causal link they refer to. In particular, we hypothesize that causal structure claims of the form "*C* causes *E*" tend to elicit the impression of a stronger causal link between *C* and *E* than causal claims of the form "*C* increases the risk of *E*" or "*C* can cause / lead to *E*".

This hypothesis is inspired by different theoretical frameworks on the semantics of causal propositions. One such framework is the mental model theory (Goldvarg & Johnson-Laird, 2001; see Johnson-Laird & Khemlani, 2017, for an overview) of causation, which attempts to explain the different meanings behind causal propositions such as " *C* causes *E*", "*C* enables *E*", or "*C* prevents *E*" (see also Wolff & Thorstad, 2017, for a different framework based on force dynamics). The key assumption of the mental model theory is that reasoners represent different causal propositions by means of different sets of mental models, where each mental model captures a particular event combination consistent with the proposition. According to the theory, the claim "*C* causes *E*" is defined by the following set of three mental models: *C* occurs and (later) *E* occurs, $\{c, e\}$, *C* does not occur and *E* does not occur, $\{\neg c, \neg e\}$, and *C* does not occur and *E* occurs (due to an alternative cause) $\{\neg c, e\}$. Since the set does not contain a mental model in which *C* occurs and *E* does not occur, $\{c, \neg e\}$, the proposition "*C* causes *E*" is assumed to capture relations in which the cause is sufficient (but not necessary) for the effect. In contrast, the proposition "*C* enables *E*" consists of one additional mental model in which *C* occurs and *E* does not occur $\{c, \neg e\}$. The proposition "*C* enables *E*" is hence assumed to describe causal relations in which a cause may fail to generate its effect. The theory does not address causal propositions such as "*C* can cause *E*" or "*C* increases the risk of *E* (happening)", but it seems plausible to assume that it would predict the mental representation of these claims to be closer to the representation of "*C* enables *E*" than the representation of "*C* causes *E*". Although we do not necessarily agree with the theory that "*C* causes *E*" implies deterministic relations, we concur that the prototypical causal relation associated with this claim is one in which *C* is relatively likely to bring about *E*.

Another reason why *modal* causal structure claims of the form "*C* can cause / lead to *E*" may elicit impressions of weaker causal strength compared to propositions like "*C* causes *E*" is that the former are more likely to prompt a *dispositional* notion of causality. Dispositional theories of causality (see, e.g., Kistler & Gnassounou, 2007; Mumford & Anjum, 2011) define causality as a relation between the properties of physical objects that bear *causal powers* (see also Waldmann & Mayrhofer, 2016, for an overview). For example, saying that sleeping pills promote fatigue according to a dispositional view means to say that sleeping pills, in virtue of being sleeping pills, possess an intrinsic power or disposition to provoke sleepiness. Importantly, dispositional theories emphasize that causal powers will often be *dormant* and become *evident* only if put into the right context, i.e., when combined in the right way with other powerful objects. Sleeping pills, for example, will not exercise their disposition unless they are ingested by an organism that possess a digestive system with the power to dissolve it. A campfire will not exercise its capacity to boil water until a jar containing water is placed in the right distance for the right amount of time. The modal verb *can* seems to emphasize the dispositional character of causal relations.

Similarly, causal propositions of the form "*C* increases the risk of *E*" may imply weaker causal relations than propositions of the form "*C* causes *E*" because the prototypical contexts in which the word "risk" is used are situations in which outcomes are probabilistic, such as in contexts of gambles or lotteries. Causal propositions of this type may be more likely to prompt a *probabilistic* reading of causality. According to probabilistic dependency theories of causality (see Williamson, 2009, for an overview) causes are probability raisers of their effects and need neither be necessary nor sufficient for their effects. The basic idea of these theories is captured by the concept of contingency, which formalizes causal probability raising by the probabilistic contrast between the probability of the effect in the presence versus the absence of the cause, $\Delta P = P(E|C,K) - P(E|\neg C,K)$, where *K* denotes the set of all relevant causal background factors.

Implications for studying causal structure induction

The idea that different terminologies used in causal structure claims may be associated with different intuitions about the strength of the underlying causal relationship is relevant for research on elemental causal structure induction. Elemental causal structure induction refers to the process by which a reasoner uses statistical data to learn whether a causal relationship exists between a single candidate cause and effect factor. In the typical paradigm used to study elemental causal structure induction (see, e.g., Griffiths & Tenenbaum, 2005;

Table 1: Causal structure claims presented in Experiment 1 together with descriptive statistics of subjects' responses.

Condition	Causal Claim	Mean ΔP (95% CI)	Median ΔP
1	"Exposure to X-ATN® causes the contraction of Lipogastrosis."	0.38 (0.15)	0.5
2	"Exposure to X-ATN® increases the risk of contracting Lipogastrosis."	0.31 (0.09)	0.33
3	"Exposure to X-ATN® can lead to the contraction of Lipogastrosis."	0.22 (0.11)	0.33
4	"Exposure to X-ATN® is related to the contraction of Lipogastrosis."	0.17 (0.14)	0.19

Lu, Yuille, Liljeholm, Cheng, & Holyoak, 2008), subjects are first presented with a fictitious cover story introducing the hypothesis that a target factor C (e.g., a medical substance, or a genetic mutation) may be the cause of another target factor E (e.g., a symptom, or a disease), then are presented with statistical data indicating a particular contingency (which is either manipulated between or within-subject), and finally are asked to indicate in some form (e.g., by reporting their degree of confidence, or by making a probability judgment) how likely the causal hypothesis is true in light of the data. Answers to these questions are assumed by recent computational models (Griffiths & Tenenbaum, 2005; Lu et al., 2008; Meder, Mayrhofer, & Waldmann, 2014) to be the result of a Bayesian inference process that computes the probability, $P(S_1|D)$ (cf. Meder et al., 2014), that the data were produced by a causal structure in which C and E are connected by a causal link (S_1: $C \rightarrow E \leftarrow A$) or by a causal structure S_0 in which the link between C and E is missing and all occurrences of the effect are due to alternative causes A (S_0: $C \quad E \leftarrow A$). What is important to note is that $P(S_1|D)$ reflects the probability of the existence of a causal link between C and E, irrespective of the strength of that link.

Interestingly, previous research on causal structure induction has focused little on the terminology used in the causal test queries subjects are asked to answer. One explanation for this may be that it has been assumed that reasoners predominantly focus on causal structure rather than on causal strength in such tasks (cf. Griffiths & Tenenbaum, 2005). The standard test query used to test causal structure induction uses formulations of the form "C causes/ makes/ produces E". For example, Lu et al. (2008) in their Experiment 3 asked their subjects after they had inspected the contingency data "How likely is it that this chemical produces headaches?". Imagine the contingency had been similar to the one depicted in Fig. 3 B, where $\Delta P = \frac{26}{72} - \frac{15}{72} = 0.21$. The effect size indicated by the data seems to be relatively small, but because it has been measured using a large sample of $N = 174$, the posterior probability of a causal link that is computed by the causal support (Griffiths & Tenenbaum, 2005) or by the structure induction model (Meder et al., 2014) is around ninety percent. Will reasoners indicate a high degree of confidence if they are asked a standard question like "How likely is it that this chemical produces headaches?". If a proposition of the form "C causes E" implies that C and E are connected by a strong causal link, it might be that subjects report lower confidence because, although representing good evidence for the existence of a causal link, the data seem to indicate that the existing link is rather weak. Compare your intuition in this case to a case where the test query would be "How likely do you think is it that C can cause E" or "How likely do you think is it that C increases the risk of E". The latter two questions seem to be more appropriate in light of the rather weak observed contingency. Whether the question is formulated in the first or the latter way seems to become less relevant, by contrast, if the observed effect size is high (cf. Fig. 3 A).

Experiment 1

The goal of Experiment 1 was to test the causal strength intuitions that are elicited by different types of causal structure claims. The different claims that we tested were inspired by the three different types of formulations that can be found in the warnings on cigarette packs. We also included a fourth statement of the form "C is related to E" because it represents a further common type of claim. The four different statements that we contrasted are shown in Tab. 1. A previous pilot study indicated that reasoners presented with these different claims tend to construct contingencies that follow the ordinal ranking in which these statements are listed in Tab. 1. The data and experimental materials of this and the second experiment can be accessed on OSF under `https://osf.io/5ngpd/`.

Methods

Participants Two hundred subjects ($M_{age} = 33.90$, $SD_{age} = 11.99$, 110 female, 87 male, 3 indicated to be neither female nor male) who were recruited via Prolific (`www.prolific.co`) participated in this online experiment and provided valid data. The inclusion criteria that we applied were a minimum

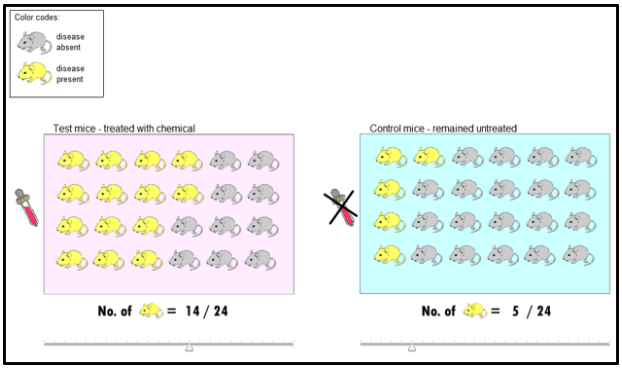

Figure 1: Illustration of the interactive animation presented in Experiment 1 that subjects used to construct the contingencies they associated with the respective causal claims. Initially, all mice were gray but when subjects moved the sliders to a particular value, the corresponding number of mice turned yellow.

age of 18 years, English as native language, and an approval rate concerning participation in previous studies of 90 percent. Prolific workers who participated in a pilot study were excluded from participation in the present experiment. We also asked subjects to participate only via laptop or desktop computer and not via smartphone or tablet, because we wanted to minimize the chances that subjects take part who are in environments (e.g., public places, subway) that might distract them. Subjects received a momentary compensation for their participation.

Design, Materials, and Procedure The study design was a 4 (causal claim: see Tab. 1; between subjects) $\times$ 2 (estimation: number of mice in the treatment group having the disease vs. number of mice in the control group having the disease; within subject) mixed design. A further balancing factor of the study will be described below.

As cover story we used a fictitious medical scenario similar to those used in previous studies on causal induction (cf. Griffiths & Tenenbaum, 2005; Stephan & Waldmann, 2018). Subjects were told that a group of biologists is investigating the effects that a particular chemical substance called X-ATN® has on the contraction of a particular disease called Lipogastrosis. Subjects read that the biologists conducted an experiment in which they examined two randomly drawn samples of mice for the disease, one sample that they had previously treated with the chemical substance and one sample that had remained untreated serving as a control group.

After having read this information about the biologists' experiment, participants proceeded to a new screen on which they were asked to assume that the biologists had finished their experiment, had analyzed the results, and had summarized their results in a short statement. Participants were then presented the biologists' conclusion, which was one of the four causal claims that are listed in Tab. 1. Then, on the same screen right below the causal statement, subjects were shown an interactive animation that looked like the illustration depicted in Fig. 1. Subjects read that the animation shows the two groups of mice before the biologists examined them for the disease. All mice were displayed in gray at this point. We then asked subjects to indicate, based on the causal statement made by the scientists, what they thought the results of the experiment most likely looked like. Subjects were prompted to use the sliders displayed in the animation below each group of mice to provide their best guess about the number of individuals that suffered from the disease in each group. When subjects clicked on and moved the sliders, the corresponding number of mice turned from gray to yellow. Whether the treatment group or the control group was displayed on the left side of the animation was counterbalanced between subjects.

After subjects had given their estimates, they proceeded to a new screen where they were asked two attention check questions, one referring to the order of the two groups in the animation (treatment group left vs. right) and the other referring to the color in which mice were depicted that had the disease. Subjects then proceeded to a new screen where they were asked to provide demographic data and to report any technical complications they may have experienced. Subjects were debriefed on the last screen of the study.

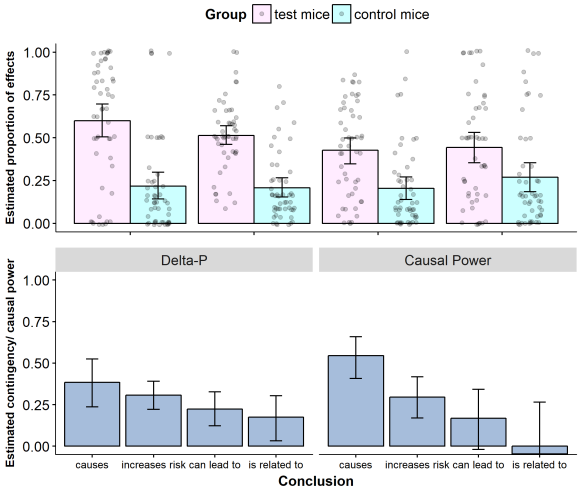

Figure 2: Results (means and 95% bootstr. CIs) of Experiment 1. The upper panel shows subjects' estimated probabilities of the effect in the presence (treatment) versus the absence (control) of the cause. Jittered dots show individuals responses. The lower left panel shows the corresponding values of ΔP, obtained from subtracting subjects' two responses. The lower right panel shows the corresponding values of causal power, obtained by dividing ΔP by $1 - P(e|\neg c)$ (Cheng, 1997).

Results and Discussion

The results are summarized in Fig. 2 and in the right part of Tab. 1. The upper graph in Fig. 2 shows subjects' slider responses transformed into probabilities. Pink and blue bars show the probabilities of the effect in the presence and the absence of the cause, respectively. The left bottom panel shows the corresponding values of ΔP, and the right panel shows the corresponding values of causal power (Cheng, 1997). The means and medians for the contingency values are listed in Tab. 1. As can be seen, subjects in all conditions tended to construct positive contingencies. That is, for all four causal statements, they assumed that the probability of the effect was higher in the presence ($M = 0.60$, $M = 0.51$, $M = 0.43$, $M = 0.44$) than in the absence of the cause ($M = 0.22$, $M = 0.21$, $M = 0.20$, $M = 0.27$). It can also be seen that the contingencies subjects constructed show a negative trend, indicating that the different types of causal statements indeed elicited different intuitions about the strength of the underlying causal relationship. As predicted, subjects in the "causes" condition tended to construct the highest contingency, although the values seem to be smaller than expected based on theories such as mental model theory. Further, it can be seen that contingencies decreased for the other statements. This trend becomes even more salient if the corresponding values of causal power are considered. Fig. 2 also shows that these contingency differences were obtained because subjects tended to construct different predictive probabilities for the different causal statements, while the base rates they constructed for the four causal claims were all similar. A poly-

nomial trend analysis conducted for the contingency values confirmed the negative linear pattern that is shown in the bottom graph in Fig. 2, $t(196) = -2.59$, $p = .01$, $r = 0.18$.

Experiment 2

Experiment 1 demonstrated that different causal structure claims tend to elicited different intuitions about the strength of that link. The goal of Experiment 2 was to test important implications for studying causal structure learning that may follow from this finding. In particular, we wanted to demonstrate that the terminology used in test queries of elemental causal-structure induction tasks is an important factor that has largely been neglected in previous studies.

Methods

Participants Two hundred and eighty subjects ($M_{age} = 33.40$, $SD_{age} = 12.24$, 173 female, 106 male, one person indicated neither female nor male) recruited via Prolific (`www.prolific.co`) participated in this study and provided valid data. The applied inclusion criteria were a minimum age of 18 years, English as native language, and an approval rate concerning participation in previous studies of 90 percent. Prolific workers who participated in Experiment 1 were excluded from participation. Subjects also were asked to participate only via laptop or desktop computer and not via smartphone or tablet, because we wanted to minimize the chances that subjects take part who are in environments (e.g., public places, subway) that might distract them. Subjects received a monetary compensation for their participation.

Design, Materials, and Procedure The study design was a 2 (contingency: $\Delta P = 0.56$ vs. $\Delta P = 0.21$) $\times$ 2 (causal query: "causes" vs. "increases risk") factorial design with both factors manipulated between subjects.

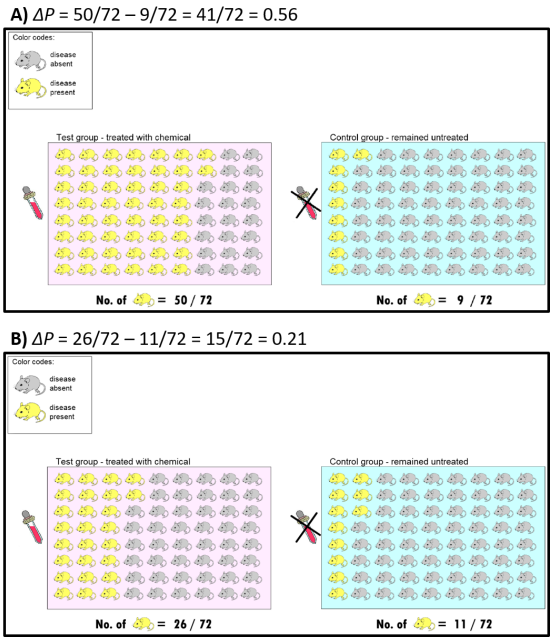

Figure 3: The two contingency data sets tested in Experiment 2. Panel A shows the data set with the high contingency and Panel B shows the data set with the low contingency.

The paradigm we used in Experiment 2 was a classical causal induction task (cf. Griffiths & Tenenbaum, 2005; Lu et al., 2008). We used the same cover story about the chemical substance and the disease as in Experiment 1. Unlike in Experiment 1, where subjects were asked to construct the results of the biologists' study, subjects in the present study were told that they will be shown the results of the biologists' experiment. We also informed participants that these results will be presented in a summary format and showed them an illustration. After subjects read the instructions, they proceeded to a new screen on which they were presented one of the two contingency data sets that are depicted in Fig. 3. A further difference from Experiment 1 was that there were 72 mice in the treatment and control group, respectively. We used this rather high sample size to obtain high values for $P(S_1|D)$ even in the $\Delta P = 0.21$ condition. The $P(S_1|D)$ values were calculated using the structure induction model of Meder et al. (2014) and are displayed by the red dashed lines in Fig. 4. Whether the treatment group of mice was displayed on the left side or the right side was counterbalanced between subjects.

The causal structure query was presented on the same screen right below the learning data. Subjects in the "causes" condition were asked a causal structure question that was formulated in the typical way found in the literature. Subjects in this condition read: "Based on the results of this experiment, how confident are you that exposure to chemical X-ATN® causes the contraction of Lipogastrosis?". Answers were provided on an eleven-point ratings scale with the endpoints labelled "It certainly does not cause the disease." and "It certainly causes the disease." (the midpoint was labelled "50:50"). The wording of the test query in the "increases risk" condition was: "Based on the results of this experiment, how confident are you that exposure to chemical X-ATN® increases the risk of contracting Lipogastrosis?". The endpoints of the scale in this case were "It certainly does not increase the risk of contracting the disease." and "It certainly increases the risk of contracting the disease.".

Results and Discussion

The results are summarized in Fig. 4. First of all, as expected and in line with previous studies testing causal structure induction (Griffiths & Tenenbaum, 2005; Lu et al., 2008), it can be seen that subjects were sensitive to contingency. They tended to report higher confidence levels in the $\Delta P = 0.56$ condition ($M = 0.84$, $SD = 0.12$ and $M = 0.88$, $SD = 0.15$ for the "causes" and "increases risk" condition, respectively) than in the $\Delta P = 0.21$ condition ($M = 0.65$, $SD = 0.17$ and $M = 0.76$, $SD = 0.18$ for the "causes" and "increases risk" condition, respectively). A 2×2 factorial ANOVA confirmed that this difference was significant, $F(1,276) = 67.52$, $p < .001$, $f = 0.50$.

Secondly, it can be seen that subjects' confidence ratings were also influenced by the causal terminology used in the test query. Overall, participants tended to report higher confidence in the existence of a causal link when asked to indicate how confident they were that the chemical "increases the

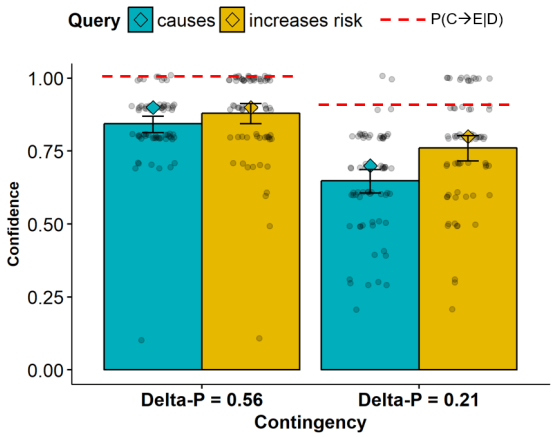

Figure 4: Results (means and 95% bootstr. CIs) of Experiment 2. Jittered dots show individuals responses. Diamonds indicate medians. Red dashed lines show the posterior probabilities of the existence of a causal link between C and E, computed based on the structure induction model by Meder et al. (2014).

risk of contracting" the disease ($M = 0.88$, $SD = 0.15$ and $M = 0.76$, $SD = 0.18$ in the $\Delta P = 0.56$ and $\Delta P = 0.21$ condition, respectively) than if they were asked to indicate their confidence that the chemical "causes the contraction" of the disease ($M = 0.84$, $SD = 0.12$ and $M = 0.65$, $SD = 0.17$ in the $\Delta P = 0.56$ and $\Delta P = 0.21$ condition, respectively). The ANOVA confirmed that this main effect was significant, $F(1,276) = 15.09$, $p < .001$, $f = 0.23$.

Thirdly, Fig. 4 shows that the effect of the causal terminology was mainly driven by subjects in the weak-contingency condition. Subjects who observed a relatively high contingency reported confidence levels that were almost equally high for both test queries. Subjects who observed a relatively weak contingency, however, were particularly reluctant to say they were confident that the chemical "causes the contraction" of the disease. However, they hesitated less saying they were confident that the chemical "increases the risk of contracting" the disease. The ANOVA confirmed a significant interaction effect between contingency and causal query, $F(1,276) = 4.07$, $p = 0.04$, $f = 0.12$. However, this interaction effect was rather small and only measured rather imprecisely in the present study. It should be replicated with higher statistical power in a future experiment.

In sum, the results of Experiment 2 indicate that subjects who are asked causal structure queries may not only indicate how much evidence the data provide for the existence of a causal link. Their confidence ratings also seem to reflect whether the data indicate a link strength that is representative for the strength implied by the terminology used in the test query. This finding has implications for the evaluation of rational models of causal structure induction. For example, the red dashed lines in Fig. 4 show the posterior probabilities of a causal link as computed by the structure induction model (Meder et al., 2014), and it can be seen that with the standard query the model fit is worse than with the alternative "increases risk" query.

General Discussion

We have shown that different types of causal structure claims also convey intuitions about the causal strength of the link they aim to communicate. This finding may have practical consequences. For example, one prediction following from it is that warning labels using phrases of the form "C causes E" will be more effective in preventing customers from consuming a particular product causing negative side effects than propositions of the form "C can cause E", because the former type of claim lets the product appear more dangerous. On the other hand, if someone wants to communicate the existence of a causal link without simultaneously eliciting a distorted representation of that link's strength, the causal structure claim should be formulated in a way conveying the right impression of the strength of the given causal relation.

We have also provided first evidence that the way causal structure queries are formulated is an important factor that should be considered in studies investigating causal structure learning. It has been argued in the literature that reasoners tend to answer causal strength queries as if they were causal structure queries (see Griffiths & Tenenbaum, 2005). Experiment 2 suggests that causal strength representations have an influence on causal structure judgments. Our finding may be particularly relevant for studies that aim to compare different models and to compare them via model fit analyses. However, it must also be said that we have only tested two types of causal structure queries and two levels of ΔP in our Experiment 2. In future experiments, larger set of contingencies, different levels of sample size, and further formulations of the test query should be tested to obtain a broader assessment of the significance of our finding.

In the present paper, we have focused on "generative" causal structure claims that postulate that a cause brings about/ produces an effect. We plan to extend our studies and to also investigate the strength intuitions elicited by "preventive" causal structure claims. Previous studies (Lu et al., 2008) suggested that "preventive" causal relationships are associated with different base rates of the effect in the absence of the target cause. Our novel contingency-construction task will be useful to investigate this effect further.

Acknowledgments This work was supported by the Deutsche Forschungsgemeinschaft (DFG) Grant WA 621/24-1.

References

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, *104*, 367–405.

Goldvarg, E., & Johnson-Laird, P. N. (2001). Naive causality: A mental model theory of causal meaning and reasoning. *Cognitive Science*, *25*, 565–610.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, *51*, 334–384.

Johnson-Laird, P. N., & Khemlani, S. S. (2017). Mental models and causation. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 147–168). Oxford University Press.

Kistler, M., & Gnassounou, B. (2007). *Dispositions and causal powers*. Aldershot, UK: Ashgate.

Lu, H., Yuille, A. L., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. *Psychological Review*, *115*, 955–982.

Meder, B., & Mayrhofer, R. (2017). Diagnostic causal reasoning with verbal information. *Cognitive Psychology*, *96*, 54–84.

Meder, B., Mayrhofer, R., & Waldmann, M. R. (2014). Structure induction in diagnostic causal reasoning. *Psychological Review*, *121*, 277–301.

Mumford, S., & Anjum, R. L. (2011). *Getting causes from powers*. New York: Oxford University Press.

Stephan, S., & Waldmann, M. R. (2018). Preemption in singular causation judgments: A computational model. *Topics in Cognitive Science*, *10*, 242–257.

Teigen, K. H., & Brun, W. (2003). Verbal expressions of uncertainty and probability. In D. Hardmann & L. Macchi (Eds.), *Thinking: Psychological perspectives on reasoning, judgment and decision making* (pp. 125–145). Wiley Online Library.

Waldmann, M. R., & Mayrhofer, R. (2016). Hybrid causal representations. In B. Ross (Ed.), *The psychology of learning and motivation* (pp. 85–127). New York: Academic Press.

Williamson, J. (2009). Probabilistic theories of causality. In H. Beebee, C. Hitchcock, & P. Menzies (Eds.), *Oxford handbook of causation* (pp. 185–212). New York: Oxford University Press.

Wolff, P., & Thorstad, R. (2017). Force dynamics. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 169–188). New York: Oxford University Press.