Answering Causal Queries about Singular Cases

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

Abstract

Queries about singular causation face two problems: It needs to be decided whether the two observed events are instantiations of a generic cause-effect relation. Second, causation needs to be distinguished from coincidence. We propose a computational model that addresses both questions. It accesses generic causal knowledge either on the individual or the group level. Moreover, the model considers the possibility of a coincidence by adopting Cheng and Novick's (2005) power PC measure of causal responsibility. This measure delivers the conditional probability that a cause is causally responsible for an effect given that both events have occurred. To take uncertainty about both the causal structure and the parameters into account we embedded the causal responsibility measure within the structure induction (SI) model developed by Meder et al. (2014). We report the results of three experiments that show that the SI model better captures the data than the power PC model.

Keywords: causal inference, generic causation, singular causation, actual causation, causal responsibility, causal attribution, Bayesian modeling

Imagine that you wake up one morning and recognize that you are haunted by a mean twinge in your head. You also know that you drank too many glasses of wine last night. Now the question arises whether your behavior last evening is causally responsible for your headache this morning. This causal query targets a *singular* instance in which one event at a specific spatio-temporal location may have caused another event that followed. The general problem with singular causal queries is that a co-occurrence of the particular events by itself does not guarantee causation. It may just be a *coincidence*. How confident can you be that this singular case of having drunk wine is the cause of your headache?

One important source that should influence our confidence is past knowledge about the contingency between drinking wine and headache. This is *generic* causal knowledge. It could either refer to cases of presence and absence of drinking and headache in an observed sample of people, or, even better with respect to the example given above, to a sample of these events in your life. The contingency in your life provides the best estimate for a generic causal relation between drinking and headache in your body.

However, knowing that there is a generic causal relation does not necessarily imply that a singular co-occurrence of the target events is causal. Unless the generic relation is deterministic, the co-occurrence may still be a coincidence. Thus, our causal judgment about singular causation must take this possibility into account.

Psychological research on causal inference adheres to two different theoretical frameworks to characterize the reasoning processes (see Waldmann, in press; Waldmann & Hagmayer, 2013; Waldmann & Mayrhofer, in press). One approach assumes that causal knowledge is grounded in knowledge about causal dependencies gleaned from observed contingencies. According to this approach, causes are *difference makers* that raise or lower the probability of an effect (e.g., Cheng, 1997; Griffiths & Tenenbaum, 2005; Meder, Mayrhofer, & Waldmann, 2014).

A different approach assumes that causal knowledge is based on a search for mechanisms and processes linking causes and effects (e.g., Ahn, Kalish, Medin, & Gelman, 1995). The two approaches need not be incompatible. Often mechanism knowledge is based on generic information about causal chains. Specific mechanisms are then simply instantiations of generic chain knowledge.

Mechanism knowledge is often not available. But even in cases in which information about intervening variables of a causal chain is in fact accessible, the question arises again how we should distinguish causation from coincidence. The joint occurrence of all elements of a chain certainly makes the possibility of a coincidence extremely unlikely, but it is in principle still a possibility. Thus, the general problem of how we should apply generic knowledge to singular cases still needs to be addressed, regardless of whether we solve this problem for a direct causal link or a causal chain.

In the present paper, we investigated how people exploit generic causal information when estimating singular causation. We here focus on cases for which it is known that both the cause event and the effect event are present, but the question needs to be answered how much confidence we can have that the co-occurrence is due to causation and not a coincidence.

Cheng and Novick (2005) have proposed a measure of *causal responsibility* that helps to answer this question. The measure is based on the assumptions of power PC theory (Cheng, 1997). It delivers the conditional probability that a cause event c is causally responsible for an effect event e given that a reasoner knows that both have occurred, $P(c \rightarrow e|c,e)$. We think that this quantity underlies judgments of singular causation when only generic information is available. One shortcoming of this measure is that it does not incorporate the reasoners' *uncertainty* concerning the underlying causal structure and the corresponding parameters (see Griffiths & Tenenbaum, 2005). Therefore, Holyoak, Lee, and Lu (2010) have proposed a causal responsibility measure that also takes parameter uncertainty into account. We go one step further here, and test a model of causal responsibility that is sensitive to both parameter uncertainty and uncertainty about the existence of the causal link between factors C and E. This

model is based on the structure induction (SI) model of diagnostic reasoning developed by Meder et al. (2014). We present three experiments in which we empirically tested the power PC theory of causal responsibility against our SI model of singular causation.

The Power PC Model of Causal Responsibility

Causal power in Cheng's (1997) theory can be understood as the probability that the target cause brings about the effect in the absence of all alternative causes of the effect. Fig. 1 (right) shows the basic causal model assumed by the power PC theory. The default assumption is that the target effect E can be independently produced by either the observed cause C with causal power (or strength) w_c or by alternative unobserved causes A with the power (or strength) w_a. It is further assumed that C and A occur independently and do not interact. Thus, C and A combine through a *noisy-OR* gate (see Griffiths & Tenenbaum, 2005; Pearl, 1988).

Based on the power PC framework, Cheng and Novick (2005) developed a measure of causal responsibility. Formally, causal responsibility is the proportion of occurrences of the effect due to the target cause C. Cheng and Novick (2005) presented formalizations of different kinds of causal responsibility. Here we are interested in how to answer the question whether an observed event c was causally responsible for an observed event e in cases in which both c and e were observed. Under the default assumptions of power PC theory, Cheng and Novick (2005) showed that this quantity is given by

$$P(c \rightarrow e|c,e) = \frac{P(c) \cdot w_c}{P(c,e)} = \frac{P(c) \cdot w_c}{P(c) \cdot P(e|c)} = \frac{w_c}{P(e|c)}, \quad (1)$$

where $P(c)$ equals b_c in Fig. 1, which denotes the base rate of C, and $P(c,e)$ denotes the joint probability of cause and effect. Since $P(c,e)$ can be rewritten as the product of $P(c)$ and the predictive probability $P(e|c)$, causal responsibility given the joint occurrence of c and e equals the causal power of c divided by the predictive probability of e given c.

The power PC theory only considers the model S_1 depicted in Fig. 1 (right) and uses maximum likelihood estimates for the parameters. It does not take into account uncertainty about the size of the parameters, nor about the causal structure (see Griffiths & Tenenbaum, 2005). Thus, it does not consider the possibility, depicted as S_0 in Fig. 1, that there is no causal arrow from C to E, meaning that all co-occurrences are due to coincidence.

A consequence of maximum likelihood estimates is that the theory predicts maximal values of causal responsibility for any contingency where the target cause is necessary (i.e., a table with an empty C cell) no matter how rarely co-occurrences of cause and effect are. Consider the two examples of such cases depicted in Fig. 2. In the left table, the effect has only occurred four times when the cause event was present. In contrast, in the right table the effect occurred sixteen times when the cause event was present. By applying the equations, one can see that the prediction of maximal responsibility is a consequence of w_c being equal to $P(e|c)$ for any table with an empty C cell.

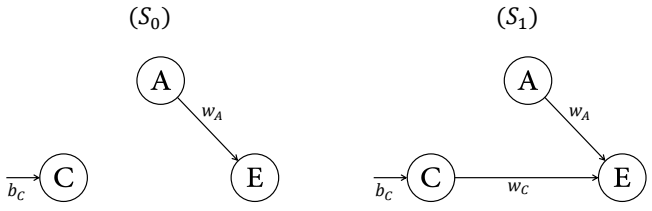

Figure 1: The two causal structures in the structure induction model. In S_0, no causal relationship exists between C and E. In S_1, C and E are causally connected. Node A represents unobservable background causes of E. The parameter b_c denotes the base rate of the cause, and w_c and w_a the causal powers of C and A, respectively.

The Structure Induction Model of Singular Causation

Contrary to the power PC theory, the structure induction (SI) model assumes that reasoners might be uncertain about both the underlying causal model and the parameters (Meder et al., 2014). Originally, the model was developed to model diagnostic inferences. Here, instead, we use the framework to model judgments about singular causation. Our key claim is that subjects assess singular causation by estimating causal responsibility within the SI model.

Causal queries about simple causal models with a single cause and a single effect are modelled by the SI theory as a Bayesian inference problem over the two mutually exclusive causal structures shown in Fig. 1 (see also Griffiths & Tenenbaum, 2005). The model formalizes the assumption that reasoners are uncertain about which of the two models underlies the data; they are also uncertain about the size of the parameters. We will briefly summarize the different computational steps of the model.

Parameter Estimation

In light of the possibility that either S_0 or S_1 might underlie the data, the model estimates the base rate and power parameters b_c, w_c, and w_a, separately for each causal structure (see Fig. 1). To express uncertainty, distributions of the parameters, rather than maximum likelihood point estimates, are inferred. According to Bayes' rule, the posterior probability distributions for the parameters of each model given the data, $P(w|D)$, is proportional to the likelihood of the data given the set of parameters w, weighted by the prior probability of the structure:

$$P(w|D) \propto P(D|w) \cdot P(w) \quad (2)$$

$P(D|w)$ denotes the likelihood of the data given the parameter values for b_c, w_c, and w_a. $P(w)$ represents the prior joint

probability of the parameters which we set to flat, uninformative *Beta*(1, 1) distributions.

Structure Estimation

The SI model separately derives the posterior probabilities for each causal structure. Applying Bayes' rule, the posterior probability for a causal structure is proportional to the likelihood of the data given the causal structure, weighted by the structure's prior probability:

$$P(S_i|D) \propto P(D|S_i) \cdot P(S_i) \tag{3}$$

$P(D|S_i)$ denotes the likelihood of the data given a particular structure, which is the integral over the likelihood function of the parameter values under the particular structure. $P(S_i)$ represents the prior probability of the structures. The model initially assumes that both structures have equal priors, i.e., $P(S_i) = 1/2$. When data are available, the posterior for a causal structure varies systematically with the observed contingency: the higher the contingency, the more likely S_1 becomes.

Causal Responsibility for Each Structure

Having estimated the parameters and the posteriors of the structures, the model computes causal responsibility separately for each parametrized structure using Equation 1. Under a noisy OR-parametrization Equation 1 can be rewritten as

$$P(c \rightarrow e|c,e) = \frac{w_c}{P(e|c)} = \frac{w_c}{w_c + w_a - w_c \cdot w_a}. \tag{4}$$

According to S_0, there is no causal connection between C and E, and any co-occurrences of c and e are coincidences. Hence, $P(c \rightarrow e|c,e) = 0$ for S_0. For S_1, Equation 4 is applied. The estimation of causal responsibility is then derived by integrating over the parameter values.

Deriving a Single Value

The final output of the model is a single estimate of causal responsibility through integrating out the causal structures by summing over the derived values of $P(c \rightarrow e|c,e;S_i)$ for each structure weighted by each structure's posterior probability:

$$P(c \rightarrow e|c,e;D) = \sum_i P(c \rightarrow e|c,e;S_i) \cdot P(S_i|D). \tag{5}$$

The incorporation of structure and parameter uncertainty by the SI model leads to systematic deviations from the predictions of the power PC model (see Fig. 2). For an illustration, consider the left contingency table depicted in (a). Although the effect never occurred in the absence of the cause, only four co-occurrences are observed. This relatively low frequency of co-occurrence is the reason for the relatively high probability of S_0. Hence, based on this information, it seems reasonable to be cautious about the causal relation of a singular co-occurrence. By contrast, consider the right table in (a). Here again, the effect never occurred in the absence of C but it occurred frequently in its presence. In this case, the data suggest a strong generic causal relationship between the factors. Now, when presented with a singular case, relatively high confidence in a case of actual causation is to be expected.

Experiment 1

The experiment tests the SI model against the power PC theory as an account of how subjects answer singular causal queries. Fig. 2 shows the predictions of the models for the two data sets that we used. The power PC model predicts invariance whereas our SI model of singular causation predicts that judgments should vary. This is because the likelihood that the observed data pattern was produced by structure S_1 varied between the two different contingencies. We also wanted to address a second question: Do subjects differentiate between generic and singular queries? To test this, we compared a generic with a singular query in the experiment.

Design, Materials, and Procedure

Sixty-one subjects completed this online experiment for monetary compensation. Fifty-six subjects (mean age = 29.16 years, *SD* = 8.4 years, 28 were female) passed our embedded attention check and were included in the analyses. The contingencies (see Fig. 2) were varied between subjects.

We used a fictitious story of a single sea devil living in an aquarium. Our subjects were asked to assume they were biologists being interested in whether noise causes the fish's antenna to flash. Subjects read they should imagine having run a single-case study with one fish to test this hypothesis. We then instructed our participants that they will see the results of this study. They read that the fish was placed in front of a loudspeaker twenty times and that the loudspeaker was off at first and activated subsequently to assess the fish's reaction.

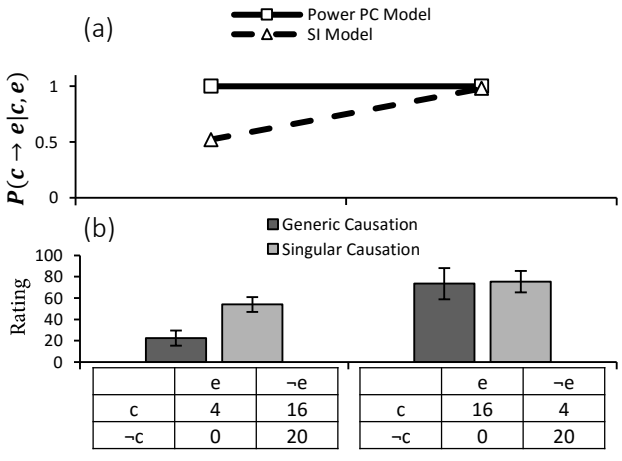

Figure 2: Predictions of the power PC and the SI model of singular causation are depicted in (a), the results in (b). Dark and light bars show means (95% CI) of the generic and singular causation ratings, respectively.

Having read the instruction, subjects saw the results of the observations arranged in a table with four columns and five rows. Below each depiction of the fish, located in front of a tiny loudspeaker, a little yellow scrip was placed that indicated the respective trial number. The initial screen showed the state of the fish as the loudspeaker was off ($\neg$ c). Next, we presented the results as the loudspeaker was activated (c), symbolized by tiny sound waves. A yellow color of the fish's flash bulb indicated the flashing of the antenna.

To test intuitions about generic causation we asked subjects how appropriate it is to say (on a rating scale from 0 to 100) that noise is a cause of the flashing of the sea devil's antenna. The corresponding question targeting singular causation asked subjects to focus on the first trial of the observations in which the fish's antenna had flashed upon noise exposure. Subjects were requested to indicate (again using a scale ranging from 0 to 100) how appropriate it is to say that the noise had caused the flashing of the fish's antenna in this particular trial. Additionally, we also asked subjects about a trial in which the antenna did not flash upon noise exposure. Furthermore, we asked them to make a predictive judgment concerning a hypothetical new trial. Here, we asked how likely it is that the fish would flash its antenna again.

Results and Discussion

The results can be seen in Table 1 and Fig. 2. Fig. 2 shows that participants judged the generic-level causal relationship between noise and antenna flashing differently for the two contingencies. On average, ratings were higher in the high contingency condition compared to the low contingency condition. As predicted by the SI model, the singular cause ratings were also different in the two conditions.

A 2 (contingency) $\times$ 4 (type of rating) mixed ANOVA with the second factor being varied within subject yielded a significant main effect of contingency, $F(1,54) = 82.00$, $p < .001$, a significant main effect of rating, $F(3,162) = 53.13$, $p < .001$, and also a significant interaction between contingency $\times$ rating, $F(3,162) = 15.74$, $p < .001$. A planned contrast comparing the singular causation ratings was also significant, $t(54) = 2.55$, $p = .01$, confirming that the ratings were higher in the high contingency condition. This difference is not predicted by the power PC model but it is predicted by the SI model.

An interaction contrast comparing the difference between the generic and the singular ratings was significant, $t(54) = 3.60$, $p < .001$. As Fig. 2 shows, this finding is due to the fact that the mean rating obtained for the singular causation question was higher than the generic causation rating in the low contingency condition. While this indicates that participants indeed conceptualized the two causal queries differently it seems that subjects answered the generic question with a causal strength rather than a structure estimate. This may also explain why there was no difference between generic and the singular ratings in the high contingency condition. As can be seen in Table 1, the additional ratings for the predictive probability were in line with the empirical predictive probability obtained in the contingency tables. Interestingly, singular ratings differed from zero in both groups for the trial in which the fish did not show a flashed antenna after noise exposure. These ratings were also higher in the high contingency condition than in the low contingency condition. This seems to reflect an unpredicted influence of generic knowledge, but it should be noted that the ratings were the lowest in the set.

In sum, the results favor our SI model of singular causation. We obtained the predicted slope of the ratings that contradicts the power PC model predictions. Furthermore, the results obtained in the low contingency condition show that subjects differentiated between a generic and a singular causal query.

Table 1. Mean ratings (*SE* of the mean) obtained for the different questions in the experiment.

	condition	
	contingency: low	contingency: high
generic causation	22.40 (3.43)	73.55 (3.43)
singular causation (e, c)	54.00 (7.12)	75.48 (4.95)
singular causation (e, $\neg$c)	10.40 (3.58)	19.03 (5.36)
predictive probability	20.40 (3.13)	76.13 (2.48)

Experiment 2

In Experiment 1 we used data sets in which the cause appeared necessary for the effect. To test our model for data sets in which all event types exist, we used in the present experiment the three contingency tables shown in Fig. 3. Again, the power PC model predicts invariance. This time it predicts values of 0.83, whereas the SI model predicts a slope. For this experiment, we also wanted to compare the SI model with a Bayesian variant of the power PC model (Holyoak et al., 2010) that takes into account parameter uncertainty but not structure uncertainty. For this model, we also used uninformative priors over the parameters. As can be seen, the predictions of this model are very similar to the predictions of the power PC model but slightly lower on average.

A second change was that the generic information did not refer to a time series of trials involving a single individual, but a sample of different individuals. Thus, we were interested in testing how subjects use generic information about a sample of different fish to respond to a causal query about a singular case.

Design, Materials, and Procedure

Thirty-eight participants (mean age = 31.23 years, *SD* = 8.90 years, 14 were female) completed this online experiment. The data sets were presented to each subject.

In this experiment, all forty observations were shown in random order on the screen in a table with five columns and eight rows. Subjects should assume they were biologists who had studied the influence of chemicals on gene mutations in populations of sea devils. They were instructed that they will see the results of these studies. Subjects also read that the

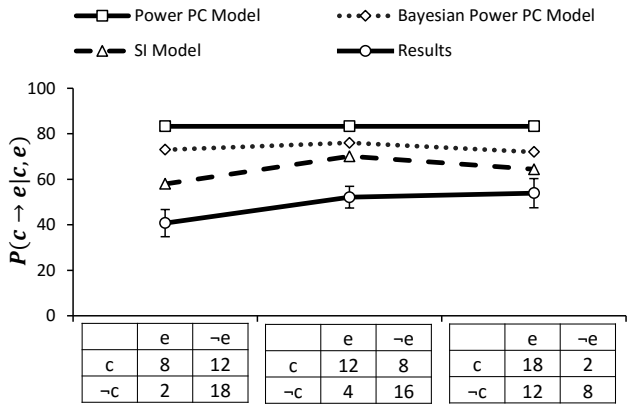

Figure 3: Model predictions and results (95% CI) for the different data sets used in Experiment 2.

chemical had been injected in half of the sample in each study and that the whole sample had later been tested for mutation. We used differently colored circles (red vs. blue) to indicate the presence vs. absence of gene mutations; a black margin around circles indicated chemical treatment. Finally, we instructed our participants that their task will be to provide a judgment concerning a single case.

For each data set, participants were asked to imagine a new individual fish which had ingested the chemical and also had the mutation. This time, we asked for the probability (on a scale from 0 to 100) that the chemical caused the gene mutation in this single case.

Results and Discussion

As can be seen in Fig. 3, ratings differed across the three data sets. A repeated-measure ANOVA with "data set" as within-subject factor was significant, $F(2,74) = 4.19$, $p = .02$. This main effect was due to the ratings for the first data set being different from those for the second and the third data set, $t(37) = 2.77$, $p < .01$. The ratings for the second and the third data set did not differ, $t(37) = 0.39$. The dampening of the upward trend across the three data sets is consistent with our SI model, although the model predicts a slight downward trend between the second and third data set. In contrast to what we observed in the first experiment, the singular causation ratings for all three data sets were lower than predicted by the SI model. One explanation for this finding may be that the type of data presentation that we used made it hard for participants to grasp the contingencies precisely, because all observations were presented on the screen in random order and also in combination with abstract symbols.

In sum, the results again favor our SI model of singular causation over the power PC model and also over a Bayesian variant of the power PC model as an account of how subjects respond to causal queries about singular causation. The present experiment further demonstrates that not only time series data but also data about samples of different individuals are used to derive a prediction for a singular case.

Experiment 3

In the first two experiments we dissociated the SI model of singular causation from the power PC model by using data sets for which the power PC model predicts invariance. To obtain additional evidence for the validity of the SI model we conducted an experiment with data sets for which the SI model predicts invariance but the power PC model does not. We obtained invariant predictions of the SI model by counteracting a slight decrease in contingency across the two contrasted data sets (see Fig. 4) with an increase of sample size.

Apart from singular causal inferences, we also tested whether participants would infer a higher probability for the existence of a generic causal relationship (i.e., for S_1) for the larger sample (Fig. 4, right), which is predicted by the SI model for generic queries. Our model predicts an interaction effect between the type of question and data set.

Design, Materials, and Procedure

All subjects saw both data sets. The type of question (confidence in generic vs. singular causation) was manipulated between subjects. 101 subjects participated in the online study (mean age = 31.40 years, *SD* = 11.10 years, 43 were female) and provided valid data.

Participants read that they should imagine to be biologists who tested the influence of the chemical "Acrinazyl" on the expression of the gene ASPM in mice in an experiment using a sample of twenty test animals. They were instructed that one half of the sample served as a control group. We presented the two halves of the sample separately on the same screen. Different colors indicated the status of ASPM. In the generic condition, we asked subjects to indicate (on a slider ranging from 0 to 100) how confident they are that the chemical does indeed raise the probability of ASPM expression. We were hoping that this test question would less ambiguously refer to generic causal relations than the one we used in Experiment 1. In the singular condition, subjects were instructed that they should imagine having picked out a single mouse from the Acrinazyl group with ASPM being expressed. They were asked how much confidence they had that it was indeed

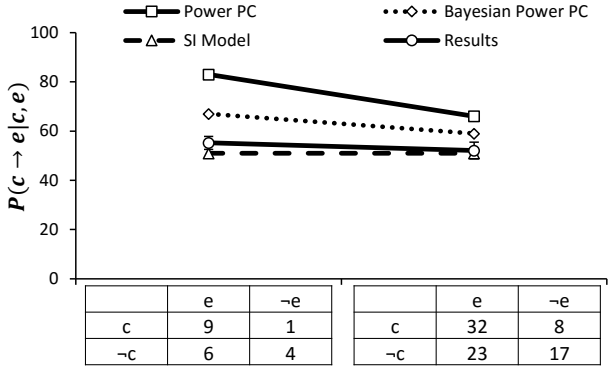

Figure 4: Model predictions and results (95% CI) for the different data sets used in Experiment 3.

the chemical that caused the gene expression in this particular mouse. Subjects indicated their confidence using a slider ranging from 0 to 100.

After participants had provided these ratings they read that they had conducted a second experiment with a larger sample size. We then showed them the results of the second study and asked them to re-assess their ratings in light of the results of this second study.

Results and Discussion

Fig. 4 shows that the singular ratings are well captured by the SI model. Moreover, as predicted by the SI model, the mean ratings for the generic causal question increased between data sets (M = 63.54, SE = 1.00; M = 68.63, SE = 1.00). A mixed ANOVA confirmed our predictions with a significant main effect for causal question, $F(1,99) = 7.42$, $p < .01$, and a significant interaction between question type and data set, $F(1,99) = 6.60$, $p = .02$. Furthermore, there was no difference between the singular causal ratings, $t(46) = 1.34$.

Overall, Experiment 3 was in line with the predictions of our SI model of singular causation. We also demonstrated again that subjects treat singular and generic causation queries differently.

General Discussion

We tested a new model of how people respond to queries about singular causation. Simply observing two consecutive events at a specific space-time location does not suffice. The question needs to be answered whether this co-occurrence manifests a causal relation or merely a coincidence. We argued that one relevant source of knowledge are generic causal relations. However, knowing, for example, that smoking generally increases the risk of lung cancer does not imply that a specific cancer patient who has smoked throughout her life has actually contracted the disease because of this risk factor. A coincidence is still possible.

One strategy that has been suggested in the literature is that judgments about singular causation should rely on unveiling causal mechanisms: observations of tar in the lung and genetic alterations in a cancer patient would strengthen the claim of singular causation. However, observing chains of singular events only helps with the question of singular causation if there is reason to assume that they are causally related. Thus, generic knowledge is necessary here as well (see also Danks, in press).

Since studying chain-like mechanisms does not fully solve the problem of how generic and singular causation are related, we here began with the simplest case with one cause and one effect event. In three experiments we showed that subjects used contingency information from both time series data of an individual and from group data when making judgments about singular causation.

We also showed that subjects differentiated between generic and singular causation. The responses to queries about singular causation were best explained by a variant of the SI model from Meder et al. (2014) that computes an estimate of causal responsibility (Cheng & Novick, 2005). Unlike its main competitors, the SI model is sensitive to the uncertainty of both the causal structures and their parameters.

Our research is just a first step. We have compared the SI model with two power PC models. However, we kept the priors uninformative and did not model different types of priors (e.g., Lu, Yuille, Liljeholm, Cheng, & Holyoak, 2008).

Another question that we want to address in future research is how subjects assess generic causal relations in time series versus group samples. Finally, it would be interesting to broaden the scope of the SI model by applying it to more complex causal models, such as causal chains, to gain insights in the important role that mechanism knowledge plays in singular causal inferences.

Acknowledgments

We thank Jonas Nagel, York Hagmayer, and Ralf Mayrhofer for helpful comments.

References

Ahn, W. K., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition*, *54*, 299–352.

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

Cheng, P. W., & Novick, L. R. (2005). Constraints and nonconstraints in causal learning: Reply to White (2005) and to Luhmann and Ahn (2005). *Psychological Review*, *112*, 694–706.

Danks, D. (in press). Singular causation. In M. R. Waldmann (Ed.), *Oxford handbook of causal reasoning.* New York: Oxford University Press.

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

Holyoak, K. J., Lee, H. S., & Lu, H. (2010). Analogical and category-based inference: A theoretical integration with Bayesian causal models. *Journal of Experimental Psychology: General*, *139*, 702–727.

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

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Francisco, CA: Morgan-Kaufmann.

Waldmann, M. R. (in press). *Oxford handbook of causal reasoning*. New York: Oxford Universtiy Press.

Waldmann, M. R., & Hagmayer, Y. (2013). Causal reasoning. In D. Reisberg (Ed.), *Oxford handbook of cognitive psychology* (pp. 733–752). New York: Oxford University Press.

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